EE
Elgar

RESEARCH HANDBOOK ON
Analytical Sociology

Edited by
Gianluca Manzo

RESEARCH HANDBOOKS IN SOCIOLOGY

© Gianluca Manzo 2021

Cover image: Hans-Peter Gauster on Unsplash.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical or photocopying, recording, or otherwise without the prior permission of the publisher.

Published by
Edward Elgar Publishing Limited
The Lypiatts
15 Lansdown Road
Cheltenham
Glos GL50 2JA
UK

Edward Elgar Publishing, Inc.
William Pratt House
9 Dewey Court
Northampton
Massachusetts 01060
USA

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

Library of Congress Control Number: 2021947840

This book is available electronically in the Elgaronline
Sociology, Social Policy and Education subject collection
http://dx.doi.org/10.4337/9781789906851

ISBN 978 1 78990 684 4 (cased)
ISBN 978 1 78990 685 1 (eBook)

Typeset by Westchester Publishing Services

Contents

List of contributors vii
Preface ix

Does analytical sociology practice what it preaches? An assessment of analytical sociology through the Merton award 1
Gianluca Manzo

PART 1 THEORETICAL CONVERSATIONS

1 Understanding the Coleman boat 49
Petri Ylikoski

2 Analytical sociology, analytic philosophy and analytical Marxism 64
Daniel Little

3 Analytical sociology and its lesser-known antecedents: structural individualism in European sociology 80
Werner Raub

4 Analytical sociology and complexity research 100
Michael Mäs

5 Analytical sociology and social stratification research 119
Carlo Barone

6 Analytical sociology and critical realism 135
Francesco Di Iorio and Francisco J. León-Medina

7 Analytical sociology and cultural sociology 155
Filippo Barbera

8 Analytical sociology and pragmatism 170
Matthew Norton

PART 2 PIVOTAL CONCEPTS AND MECHANISMS

9 Action 186
Clemens Kroneberg and Andreas Tutic

10 Interactions 204
Michael Mäs

11 Structure 220
Emily Erikson and Eric Feltham

Gianluca Manzo - 9781789906851

12 Norms 233
Pierre Demeulenaere

13 Coupled contexts 249
Filippo Barbera and Nicola Negri

14 Contingency and randomness: a modal approach 264
Ivan Ermakoff

15 Cumulative advantage 286
Freda B. Lynn and Hannah W. Espy

16 Network amplification 308
Filiz Garip and Mario D. Molina

17 Complex contagions 321
Damon Centola

PART 3 PIVOTAL METHODS

18 Archival data 337
Katharina Burgdorf and Henning Hillmann

19 Digital data and methods 352
Viktoria Spaiser

20 Advances in mediation analysis 364
Kinga Makovi and Christopher Winship

21 Experiments 392
Giuseppe A. Veltri

22 Game-theoretic models 414
Wojtek Przepiorka

23 Network mechanisms and network models 432
Christoph Stadtfeld and Viviana Amati

24 Agent-based computational models 453
Andreas Flache and Carlos A. de Matos Fernandes

25 The many model approach 474
Scott E. Page

Coda—The past and future of analytical sociology 490
Peter Hedström

Index 506

Gianluca Manzo - 9781789906851

Contributors

Viviana Amati is a postdoctoral researcher at ETH Zürich, Switzerland.

Filippo Barbera is a professor of economic sociology at the University of Turin and Fellow at the Collegio Carlo Alberto (Turin), Italy.

Carlo Barone is a professor of sociology at the Paris Institute of Political Studies (Sciences Po), France.

Katharina Burgdorf is a PhD candidate in sociology at the University of Mannheim, Germany.

Damon Centola is the Elihu Katz professor of communication, sociology and engineering at the University of Pennsylvania, USA.

Carlos A. de Matos Fernandes is a PhD candidate in sociology at the University of Groningen, the Netherlands.

Pierre Demeulenaere is a professor of sociology at Sorbonne University, France.

Francesco Di Iorio is an associate professor of philosophy at Nankai University in Tianjin, China.

Emily Erikson is an associate professor of sociology at Yale University, USA.

Ivan Ermakoff is the Sewell-Bascom professor of sociology at the University of Wisconsin-Madison, USA.

Hannah W. Espy is a PhD candidate in sociology at the University of Iowa, USA.

Eric Feltham is a PhD candidate in sociology at Yale University, USA

Andreas Flache is a professor of sociology at the University of Groningen and the Interuniversity Center for Social Science Theory and Methodology (ICS), the Netherlands.

Filiz Garip is a professor of sociology and public affairs at Princeton University, USA.

Peter Hedström is a professor of analytical sociology and the founder of the Institute for Analytical Sociology at the University of Linköping, Sweden.

Henning Hillmann is a professor of economic and organizational sociology at the University of Mannheim, Germany.

Clemens Kroneberg is a professor of sociology at the University of Cologne, Germany.

Francisco J. León-Medina is a professor of sociology at the University of A Coruña, Spain.

Daniel Little is a professor of philosophy at the University of Michigan-Dearborn, professor of sociology and public policy (by courtesy) at the University of Michigan, Ann Arbor, and chancellor emeritus of the University of Michigan-Dearborn, USA.

Gianluca Manzo - 9781789906851

Freda B. Lynn is an associate professor of sociology at the University of Iowa, USA.

Kinga Makovi is an assistant professor of social research and public policy at New York University, Abu Dhabi, United Arab Emirates.

Gianluca Manzo is a professor of sociology at Sorbonne University, France.

Michael Mäs is a professor of sociology at the Karlsruhe Institute of Technology, Germany.

Mario D. Molina is a postdoctoral associate in the Social Science Division at New York University, Abu Dhabi, United Arab Emirates.

Nicola Negri was a professor of economic sociology at the University of Turin, Italy.

Matthew Norton is an associate professor of sociology at the University of Oregon, USA.

Scott E. Page is the John Seely Brown Distinguished University Professor of Complexity, Social Science, and Management at the University of Michigan, USA.

Wojtek Przepiorka is an associate professor of sociology at Utrecht University and the Interuniversity Center for Social Science Theory and Methodology (ICS), the Netherlands.

Werner Raub is an emeritus professor of sociology at Utrecht University and the Interuniversity Center for Social Science Theory and Methodology (ICS), the Netherlands.

Viktoria Spaiser is an associate professor in sustainability research and computational social science at the University of Leeds, UK.

Christoph Stadtfeld is an associate professor of social networks at ETH Zürich, Switzerland.

Andreas Tutic is a Heisenberg-Fellow of the German Research Foundation (DFG) at the University of Leipzig, Germany.

Giuseppe A. Veltri is a professor of research methodology and cognitive sociology at the University of Trento, Italy.

Christopher Winship is the Diker-Tishman professor of sociology at Harvard University and a member of the senior faculty at the Harvard Kennedy School of Government, USA.

Petri Ylikoski is a professor of science and technology studies (sociology) at the University of Helsinki, Finland.

Gianluca Manzo - 9781789906851

Preface

The principles and the methods of analytical sociology have been under elaboration and critical debate over the last twenty-five years. After a period of sustained programmatic discussion, substantive works inspired by the analytical sociology's research strategies have started to appear in the major journals of the discipline. A prize recognizing the best article in analytical sociology was created in 2012 and, since 2013, it has been presented each year at the annual meeting of the *International Network of Analytical Sociology*.

The present *Research Handbook on Analytical Sociology* intends to contribute to the further development of analytical sociology's content and frontiers. It builds on the shoulders of giants. In particular, it is thought as a follow-up on the foundational *Oxford Handbook of Analytical Sociology*, edited by Peter Bearman and Peter Hedström in 2009. It has the ambition to provide an updated pedagogical resource for readers that are new, or only partially acquainted, to analytical sociology. This new *Research Handbook* is animated by three objectives.

First, it wants to establish a more explicit dialogue between analytical sociology and potentially close, as well as more distant, research fields and theoretical perspectives. This is the common goal of the eight chapters contained in the *Research Handbook*'s Part 1, titled "Theoretical conversations". Second, it wants to provide an updated overview of the main theoretical resources that, for now, can be considered the basic building blocks for designing models of social mechanisms. This is the common goal of the nine chapters contained in the *Research Handbook*'s Part 2, titled "Pivotal concepts and mechanisms". Third, it wants to foster a new form of methodological pluralism under constraint for analytical sociologists; the constraint being the awareness that different methods have different potentialities with respect to the formalization and the empirical test of the different parts of a model of a mechanism. This is the common goal of the eight chapters contained in the *Research Handbook*'s Part 3, titled "Pivotal methods".

The present *Research Handbook* is primarily intended to speak to readers that are new to analytical sociology. Among them, advanced under-graduate and doctoral students, as well as young scholars in sociology, economics, demography, social psychology or computer science, are a special target. They may have heard about analytical sociology without exactly knowing the history, the theoretical literature and debates, the main concepts, and the methods of analytical sociology. With this *Research Handbook*, my hope is that these newcomers can access a new synthesis of the relevant theoretical, conceptual and methodological resources to understand what analytical sociology is and how it can be practiced in an effective way. The goal is to make the reader capable of navigating the diversity of options that are present in the field of analytical sociology, thus making her/him fully aware of what a research design inspired from the analytical perspective could look like. For this reason, contributors were asked to conceive their chapters as critical overviews of existing problems, points of view, evidence and technical options on a given theoretical perspective, concept, mechanism or method. Fundamentally, this new *Research Handbook* is intended to be practically useful to do research *with* analytical sociology rather than

Gianluca Manzo - 9781789906851

providing another programmatic "polished" and "policed" *manifesto* for believers in analytical sociology.

With this aim in mind, the *Research Handbook* is introduced by a general chapter where the book's editor overviews the progressive evolution of analytical sociology on a programmatic level over the last twenty-five years and then systematically describes what analytical sociology looks like in substantive works that are supposed to be first-class illustrations of analytical sociology's research *agenda*. This analysis confirms that, to a large extent, analytical sociology exists in practice, not only in principle. However, the substantive works investigated also reveal that several aspects of the way analytical sociology is put to work can be improved. In particular, "practiced" analytical sociology seems hesitant to exchange with analytical sociology's theoretical *manifestos*; mechanism comparison and accumulation is still under-developed, and the specific contribution of each method for mechanism model-building and testing is often left unspecified.

These three problems motivate the content of the three main parts in which the *Research Handbook* is organized. In particular, the chapters included in the first part of the book, "Theoretical conversations", aim at the further clarification of analytical sociology's principles and paradigmatic antecedents and borders. This should help to make it clearer when one should refer to analytical sociology in her/his substantive works as well as to facilitate the dialogue between analytical sociology and other perspectives that one may want to mobilize when addressing a particular substantive problem.

This operation of paradigmatic clarification is pursued in various, but complementary, ways. First, Chapter 1 by Petri Ylikoski invites the reader to go back to a fundamental conceptual tool of analytical sociology, i.e. the Coleman boat, and shows that the flexibility of this tool for theory construction goes well beyond the usual interpretations of the Coleman's diagram. Second, Chapter 2 by Daniel Little and Chapter 3 by Werner Raub reconsider some aspects of the historical roots of analytical sociology; in particular, analytical Marxism and analytic philosophy on the one hand, and the research program of "structural individualism" on the other hand. Third, Chapter 4 by Michael Mäs and Chapter 5 by Carlo Barone investigate similarities and differences between analytical sociology and potentially close research fields like complexity science and social stratification research respectively. Finally, a more explicit comparison between analytical sociology and theoretical perspectives that are in principle distant from analytical sociology is proposed by Francesco Di Iorio and Francisco J. León-Medina in Chapter 6 on critical realism, by Filippo Barbera in Chapter 7 on cultural sociology, and by Matthew Norton in Chapter 8 on pragmatist-inspired sociology.

The chapters included in the second part of the book "Pivotal concepts and mechanisms" have the common goal of providing an updated overview of the theoretical resources that can be considered the basic building blocks for designing models of social mechanisms thus, hopefully, helping analytical sociology to increase knowledge accumulation and to favor mechanism comparison.

In particular, the first four chapters of this second part present conceptual building blocks at various levels of abstraction. From the most microscopic to the most macroscopic: Chapter 9 by Clemens Kroneberg and Andreas Tutic covers various understandings of action theory within analytical sociology; Chapter 10 by Michael Mäs treats dyadic interactions; Chapter 11 by Emily Erikson and Eric Feltham presents the concept of structure, and more particularly its

Gianluca Manzo - 9781789906851

understanding in terms of network structures; Chapter 12 by Pierre Demeulenaere contains an analytical treatment of norms; and Chapter 13 by Filippo Barbera and Nicola Negri discusses various types of larger contexts, among which are organizational settings. Generic elements shared by all these basic building blocks, like contingency and randomness, are then discussed by Ivan Ermakoff in Chapter 14. The last three chapters of this second part systematize the knowledge accumulated so far on generic mechanisms that have been especially studied by analytical sociologists. This is the case of cumulative advantage mechanisms covered by Freda B. Lynn and Hannah W. Espy in Chapter 15; network amplification studied by Filiz Garip and Mario D. Molina in Chapter 16; and complex contagions fueled by particular features of social networks reviewed by Damon Centola in Chapter 17.

Finally, the chapters included in the third part of the book, "Pivotal methods", move on to data and methods and have the common goal of increasing analytical sociologists' awareness of how various types of data and methods can specifically contribute to formalize and document empirically various aspects of a model of a social mechanism thus, hopefully, helping analytical sociology to increase even more the transparency of the proposed explanations.

In particular, the first two chapters of this third part give priority to information that have been less frequently exploited by analytical sociologists so far. This is the case of historical and qualitative, as well as numeric, archival data discussed by Katharina Burgdorf and Henning Hillmann in Chapter 18; this is also the case of digital data covered by Viktoria Spaiser in Chapter 19. In spite of their focus on data and methods, these two chapters, in fact, also contribute to conversations between analytical sociology and other research fields, namely network-oriented historical sociology (in the former case) and computational social sciences (in the latter case). The following five chapters cover methods, placing a strong emphasis on causal inference such as mediation analysis, treated by Kinga Makovi and Christopher Winship in Chapter 20, and experiments covered by Giuseppe A. Veltri in Chapter 21, as well as methods that focus on dynamic processes and generativity like game theory, presented by Wojtek Przepiorka in Chapter 22 (in connection with social dilemmas), network models discussed by Christoph Stadtfeld and Viviana Amati in Chapter 23, and agent-based computational models and simulations problematized by Andreas Flache and Carlos A. de Matos Fernandes in Chapter 24. To emphasize without ambiguity the complementarity of these resources to represent and document a given social mechanism, Scott E. Page discusses in Chapter 25 the relevance for analytical sociology of his claim on the many-model approach.

Through a personal and provocative outlook on the past and the current state of analytical sociology, Peter Hedström's *Coda* closes the *Research Handbook* by inviting the reader to reflect on what action theory, models of social interactions and disciplinary alliances can do to optimally foster the further development of analytical sociology.

The homogeneity of edited volumes should never be claimed without precaution. In the present case, perfect consistency was not even an objective. Analytical sociology is still under construction. Variations on a common theme should be considered a richness of the approach. The present *Research Handbook* did not want to polish asperities. Despite this, it is surprising to see a *posteriori* how strongly an impression of family resemblance arises when the theoretical, conceptual and methodological contributions are considered together. I do hope that readers will have the same feeling, and benefit from this collective effort.

Gianluca Manzo - 9781789906851

Contributors should all be praised for playing seriously a game, whose rewards are highly uncertain. I am deeply grateful to them for their commitment to this project. I am especially in debt to Viviana Amati, Filippo Barbera, Andreas Flache, Peter Hedström and Werner Raub for reading various parts of the manuscript and providing helpful written remarks. I would also like to thank Hans-Peter Blossfeld for giving me the responsibility for editing a new *Research Handbook* on analytical sociology, and Daniel Mather at Edward Elgar Publishing for his continuous editorial support and assistance. The usual disclaimers apply.

Gianluca Manzo - 9781789906851

Does analytical sociology practice what it preaches? An assessment of analytical sociology through the Merton award

Gianluca Manzo[1]

INTRODUCTION

Mechanism-based explanations are at heart of analytical sociology. If one considers Hedström and Swedberg's article, published in *Acta Sociologica* in 1996, as the beginning of the explicit and systematic elaboration of a research program in sociology based on the concept of social mechanism, the field of analytical sociology has been under debate for about twenty-five years. This time is sufficiently long to appreciate, on the one hand, how the principles and methods of analytical sociology have evolved on a programmatic level, while on the other hand questioning whether analytical sociology practices what it preaches. This second part of the exercise seems particularly urgent. Some have argued that, while "the proof of the pudding is always in the eating", analytical sociology spends more time discussing epistemological and meta-theoretical issues rather than applying its programmatic principles and methods to substantive problems (Edling and Rydgren 2016: 1139).

In this chapter, I first synthetically reconstruct the progressive evolution of analytical sociology on a meta-theoretical level and show that analytical sociology is currently predicated under various programs and sub-programs (section 1). Then, with the aim to assess what analytical sociology has delivered so far in practice, I take these programmatic variations as a benchmark and compare them against a specific set of journal articles, namely those having received the Merton award (see Table I.1), the prize that the *International Network of Analytical Sociology* first gave in 2013 in recognition of the best article on analytical sociology.

In particular, I will go through five steps. After preliminarily describing some formal signs of the Merton awarded article's academic visibility (section 2), I first investigate the extent to which these articles explicitly referred to analytical sociology and mechanism-based explanations to frame their substantive investigations (section 3). Second, I investigate whether the Merton awarded articles conformed to analytical sociology's requirement of focusing on well-defined macroscopic patterns/dynamics with a clear explanatory goal in mind (section 4). Third, I reconstruct the explanatory machinery of each Merton awarded article and assess the extent to which it conformed to analytical sociology's requirement of explaining by formulating well-specified models of mechanisms, thus also assessing how the concept of mechanism is understood within these works (section 5). Fourth, with the specific aim in mind of establishing the extent to which particular methods dominate analytical sociology's empirical works, I investigate how the Merton awarded articles combined data and methods to prove

[1] I would like to thank Viviana Amati, Andreas Flache, and Werner Raub for providing helpful written comments on an early version of this chapter. The usual disclaimers apply.

Gianluca Manzo - 9781789906851

the empirical plausibility of the proposed mechanisms (section 6). Finally, with some critiques of analytical sociology in mind, I consider again the explanatory mechanism formulated by each Merton awarded article and investigate whether actors' actions are framed within rational choice theories (section 7.1) as well as the extent to which the proposed mechanisms are designed by using "anomalous" elements, i.e. concepts, theories, or broader research perspectives that analytical sociology's *manifestos* either criticized or did not explicitly include within the analytical sociology's basic toolbox (section 7.2).

The most general result of this analysis was that, although the authors of the Merton awarded articles rarely referred to analytical sociology's *manifestos* explicitly, they applied in practice all the theoretical principles and, to a large extent, the methodological principles too; those *manifestos* progressively codified with increasing explicitness. The analysis also revealed that analytical sociology's substantive investigations are less engaged with rational choice theories, but are characterized by a higher level of theoretical pluralism than analytical sociology's critics are usually ready to admit.

In conclusion, I build on these results (summarized in section 8) to suggest five guidelines for the further development of analytical sociology and explain how the present *Research Handbook* intends to contribute to this process.

1. PROGRAMS AND SUB-PROGRAMS IN ANALYTICAL SOCIOLOGY

Over the last twenty-five years, the principles and the methods of analytical sociology were progressively clarified and codified. The perspective moved from a set of general statements on how to increase the explanatory value of sociological theory to a research strategy clearly articulated around a well-defined sequence of operational steps (section 1.1). This founding stage raised a fair amount of critiques (section 1.2), which led to a new formulation of the analytical sociology's research core, as well as to the awareness that variations now exist in this core (section 1.3). The presence of programs and sub-programs of analytical sociology makes it important to assess empirically how analytical sociology materialized in substantive works that were delivered independently from edited volumes on this perspective (section 1.4).

1.1 The Founding Stage (1996–2009)

The very first agenda-setting article where the concept of social mechanism was systematically treated (Hedström and Swedberg 1996) did not even employ the term "analytical sociology". It essentially defended the general claim that well-specified models of mechanisms could help sociological theorizing to escape mere conceptual re-labelling and story-telling and simple descriptions of correlations among variables. Mechanisms were seen as multi-layered chains of events that, by connecting actors' opportunities, desires, and beliefs to the social context in which actors are embedded, dynamically led to (associations of) events at a higher level of abstraction (ibid., 296–8).

Hedström and Swedberg's (1998a) subsequent edited volume demonstrated that this message had the potential to rally a variety of scholars around a common research program, but the specific content of this research program remained relatively unspecified. The term "analytical sociology" only appeared once in Hedström and Swedberg's (1998b, 15) introductory essay.

Gianluca Manzo - 9781789906851

The "analytical" approach to social theory they defended essentially consisted of (1) a general claim on the explanatory value of building abstract models of social mechanisms (ibid., 7–11), and (2) a defense of an open version of methodological individualism (ibid., 11–13), and a critique of variable-based sociology (ibid., 15–17). To add concreteness, Hedström and Swedberg (1996: 293–6; 1998b: 17–21) discussed several pieces of research illustrating, in their eyes, the distinctive features of mechanism-based theorizing, namely Merton's self-fulfilling prophecy, Granovetter's threshold models, and Coleman's study of the diffusion of innovations through social networks. But this did not translate into a specific and structured research protocol that could have been used to build models of mechanisms, formalize them, and investigate them empirically.

Hedström (2005) constituted an important step into this direction. In this book, the idea of mechanism-based explanation is no longer the focus of the investigation: it became one of the components of "analytical sociology"; in particular, its favorite understanding of explanation (ibid., 2). Analytical sociology is then presented as a set of principles and tools for building this type of explanations and connecting them to empirics.

More particularly, as to principles, Hedström (2005) introduced three new elements: (a) an invitation to balance the abstraction of models of mechanisms with the realism of the assumptions on which they are based (ibid., 3, 62, 65); (b) the requirement of systematically relating the basic components of the individual actions (in particular, beliefs, desires, and opportunities) to the structure of the interactions in which actors are embedded (ibid., 42–56); (c) the imperative of not restricting the hypotheses concerning individuals' action to cost-benefit analysis, utility maximization or optimality reasoning (ibid., 41, 60–6). As to tools, three proposals were made: (a) social networks analysis was seen as a way to describe, on the one hand, aggregate patterns of explanatory interest (ibid., 5, 67) and, on the other hand, specific pieces of more complex models of mechanisms, in particular the relevant interaction structures that may constraint actors' actions (ibid., 76f, 93–8); (b) numerical simulations of agent-based computational models were proposed as the most flexible tool to study the micro-to-macro-transition part of a model of a mechanism, in particular when actors' choices are contingent on interactions with other actors (ibid., 76–87); (c) a specific type of data-driven agent-based computational simulations—where the agents' behavior is based on previously estimated statistical models—were also proposed as a pivotal tool to reach empirically plausible simulated models of social mechanisms (ibid., chapter 6).

However, one must wait for Hedström and Bearman (2009) to see these points translated into a synthetic bullet-pointed research recipe for the first time. After discussing Schelling's segregation model (ibid., 12–13) and more recent works on network topologies formation (ibid., 14–15) as illustrations of analytical sociology's principles and methods, Hedström and Bearman (2009: 16) codified the bulk of the "explanatory strategy" of analytical sociology in the following terms:

1. We start with a clearly delineated social fact that is to be explained.
2. We formulate different hypotheses about relevant micro-level mechanisms.
3. We translate the theoretical hypotheses into computational models.
4. We simulate the models to derive the type of social facts that each micro-level mechanism brings about.
5. We compare the social facts generated by each model with the actually observed outcomes.

Hedström and Bearman (2009: 16) labeled this research strategy "generative methodology" and claimed that "[it] is generally applicable and it is a crucial part of the analytical toolbox".

Gianluca Manzo - 9781789906851

1.2 The Critical Effervescence (2007–2014)

Interestingly, the more clearly analytical sociology's core was specified, the more explicitly and virulently it was criticized. The epistemological, theoretical, and conceptual aspects of this core were abundantly discussed. The concept of mechanism was seen as (a) poorly defined (Gross 2009: 360–2); (b) the epistemological status of a mechanism was judged unclear (Mayntz 2004: 239; Reiss 2007: 166); (c) mechanism-based explanations were considered as necessarily incomplete (Opp 2005: 169; Steel, 2004: 61–4); (d) the opposition between mechanism-based explanations and the covering-law model of explanation were regarded as unconvincing (Opp 2005: 174–7; 2007: 117–18; Norkus 2005: 352–5); (e) analytical sociology's theory of action was judged as substantially equivalent to rational choice theory (Abbott 2007a; Gross 2009: 365–7; Opp 2013). In contrast, the methodological components of the analytical sociology's research program received only rapid comments. In particular, some objected that analytical sociology underestimated the role of multivariate statistical methods (Opp 2007: 121); others observed that the concept of mechanism in fact is ultimately reducible to that of variable (Morgan 2005: 31); still others regarded the emphasis on computer simulations as excessive (Abbott 2007b: 1; Sawyer 2007: 260).[2]

These commentators shared a common goal. They wanted to question the distinctiveness of analytical sociology. Some were more explicit than others. For instance, Lizardo (2012) claimed that analytical sociology's emphasis on agent-based computer simulations made this perspective soluble within computational social sciences; Little (2012) regarded analytical sociology's emphasis on micro foundation as a sign that analytical sociology was no more than a new brand name for the doctrine of methodological individualism (for a reply, see Barbera 2012; Manzo 2012a). Surprisingly enough, this same view of "the handbooks of analytical sociology" was expressed by Boudon (2012) himself, whom Hedström (2005: 6, 7, 23, 112–3, 145) regarded as one of the pioneers of analytical sociology (for a reply, see Manzo 2012b: 50–7). Demeulenaere (2011: 5) also argued that the main contribution of analytical sociology for contemporary social sciences consisted in avoiding usual "misconceptions" about methodological individualism.

Analytical sociologists recognized the importance of these critiques but they also noted that they were selective. Critics typically focused on one, or a few, theoretical principles of analytical sociology and, when methodological elements were considered, only a few aspects were evoked. For instance, Winship and Morgan (2014: 341, fn. 15) expressed skepticism against "computer simulations" in general, but they did not discuss in detail how analytical sociology proposed to connect them to data. By doing so, analytical sociology's overall picture was lost. The potential benefits of the progressive codification of analytical sociology's core vanished. The analytical sociology's project was integrative and inclusive. Its distinctiveness was to find, in the set of epistemological, theoretical, and methodological elements, that the analytical sociology's *manifestos* progressively clarified. Instead, the originality evaporated as soon as the parts were isolated from the whole (see Manzo 2011).

---

[2] A systematic discussion of these objections can be found in Manzo (2007; 2010: 140ft, 142–57; 2013; 2014: 15–16) and Ylikoski (2013).

Gianluca Manzo - 9781789906851

1.3 A Re-Formulation and New Reactions (2014–)

To make this point explicit, Manzo (2014: 7–9) further elaborated analytical sociology's principles on the basis of the critiques accumulated over the years and proposed to combine analytical sociology's core meta-theoretical and methodological ingredients within a synthetic set of seven postulates (P):

> P1: use concepts that are as clear and precise as possible to describe both the facts to be explained and the explanatory hypotheses/facts mobilized to explain them, while avoiding all linguistic obscurity and convolutedness;
> P2: mobilize the best quantitative and qualitative empirical information available and use the technical tools best suited to describing the facts to be explained;
> P3: in order to explain the social outcome(s) described, first formulate a "generative model," that is, a model of a (set of) mechanism(s), where a mechanism is a set of entities and activities likely to trigger a sequence of events (i.e., a process) likely to bring about the outcome(s);
> P4: in order to formulate the "generative model," provide a realistic description of the relevant micro-level entities (P4a) and activities (P4b) assumed to be at work, as well as of the structural interdependencies (P4c) in which these entities are embedded and their activities unfold;
> P5: in order rigorously to assess the internal consistency of the "generative model" and to determine its high-level consequences, translate the "generative model" into an agent-based computational model;
> P6: in order to assess the generative sufficiency of the mechanisms postulated, compare the agent-based computational model's high-level consequences with the empirical description of the facts to be explained;
> P7: in order to prove that the hypothesized micro- and network-level assumptions are not only generative sufficient but also empirically grounded, inject as much individual- and relational-level quantitative, qualitative, and/or experimental data as possible into the agent-based computational model and re-analyze its behavior and high-level consequences.

Manzo (2014: 29–34) carefully explained why agent-based computational models should be especially considered for the formal deduction of the macroscopic consequences of a model of a mechanism but, at the same time, he explicitly recognized that P1–P7 should not be seen as a rigid research sequence one must go through in any single work (ibid., 10). The proposed list of theoretical and methodological options should instead be interpreted as a logically well-organized set of "negative heuristics" telling us what research path should be avoided and of "positive heuristics" telling us what paths should be pursued. The *sui generis* nature of analytical sociology, Manzo argued, is to advise us to keep in mind the entire set of "heuristics" even if, for a variety of practical reasons, it may not be possible, or even not necessary, to perform all the requisite operations within a given piece of research (ibid., 10).

Manzo's re-formulation of analytical sociology's principles and methods had the interest of forcing scholars to tell explicitly what postulates they are ready to accept as foundational for analytical sociology. As a by-product, it became apparent that analytical sociology is now predicated under different programmatic regimes, depending on what principles are the most valued by a given author.

In particular, Kalter and Kroneberg (2014: 107) recognized that Manzo's set of principles:

> […] allows us to clarify, again, the common denominator of explanatory and analytical sociology, as scholars in both traditions generally agree on principles P1–P4: emphasising clarity and precision, carefully establishing the explanandum, and formulating a generative model of the underlying mecha-

Gianluca Manzo - 9781789906851

> nisms along the lines of structural individualism", but they advised not to give a "truly foundational status" to methodological principles "[…] P5–P6 […] as they are all about agent-based models and their empirical testing and calibration […]".

Kalter and Kroneberg (2014: 108) preferred instead not to restrict programmatically the way of formalizing and testing generative models of social mechanisms to any particular methods or research design. Similarly, Bouvier (2018: 19) observed that:

> […] Manzo highlights a very specific sub-program of analytical sociology (p. 6), which explicitly emphasizes the role of agent-based computational modeling (pp. 7–10) and, generally speaking, formal modeling (in particular graph theory).

Bouvier did not contest "[…] the legitimacy of such a sub-program […]", but he recommended to put more emphasis on "the issues of rationality and intentionality", thus valuing analytical sociology's general contribution to action theory more than its empirical application through specific methodologies.

With respect to action theory itself, Kroneberg and Tutic (this volume) also recognized the existence within analytical sociology of a "weak" and a "strong program", the former being agnostic about the action theory that should inform theoretical model building (as illustrated by Manzo's postulate 4, which is silent about how to depict the behavior of the micro-level entities under exam), whereas the latter still believed in the importance of seeking a general theory of action to be used in an axiomatic way (see also Hedström this volume).

Keuschnigg et al.'s (2018) recent analysis of the potential relationships between analytical sociology and the larger field of computational social sciences also contains traces of the co-existence within current analytical sociology of a general program focusing on generic meta-theoretical principles with a specific sub-program emphasizing particular methods to connect theory and empirics. In particular, Keuschnigg and colleagues described "[…] the core defining characteristics of AS […] under three headings related to the importance of explanation, the modeling of dynamic social processes, and realism […]" (ibid., 4), and, when giving details on the second elements, they noted that:

> the guiding idea of AS is that we explain collective or aggregate outcomes: (1) by developing a model—*possibly* a simulation model—that shows how individuals can bring about such an outcome, and (2) by empirically demonstrating that the proposed model indeed is plausible in the specific case. (ibid., 5; emphasis added)

Thus, again, computer simulations are presented as an ideal goal for analytical sociology, not as a necessary ingredient.[3]

---

[3] It should also be noted that Keuschnigg et al. (2018: 8–10) argued that the added value of analytical sociology for computational social sciences consists in its emphasis on building explanatory models of mechanisms rather than on tools. The latter—among which they included agent-based simulations, macro-oriented experiments, counterfactual approaches to causality, and inductive analyses using machine learning techniques on more or less unstructured data—are brought to analytical sociology from computational social sciences (ibid., 6–8, 11).

Gianluca Manzo - 9781789906851

1.4 Programs and Sub-Programs in Substantive Works

But how were these programmatic variations materialized in substantive works inspired by analytical sociology? Some regretted that the movement of progressive codification of analytical sociology's principles and methods was disconnected from their applications. Edling and Rydgren (2016: 1139) overtly denounced this limitation, and claimed

> […] the literature on social mechanisms is still too preoccupied with intratheoretical and metatheoretical discussions, and we find very few empirical applications […]; the proof of the pudding is always in the eating and we strongly believe that examples of good empirical research, not theoretical manifestos, are what drive the evolution of social science.

To advance in that direction, Edling and Rydgren (2016: 1141–4) collected six mechanism-oriented pieces of empirical research, compared them, and concluded that, although these pieces did not share the same "definition of what a mechanism actually is" (ibid., 1140), they illustrated four fundamental points of the social mechanism approach: (1) they all started with a research question that "identify a real-world phenomenon and ask a pertinent question that demands explanation"; (2) they identified mechanisms at different levels all involving some form of micro foundation and reference to social interactions; (3) the proposed mechanisms were not exclusively associated to "one particular model of man the way rational choice theory does" (ibid., 1140); (4) various types of data were used to show the explanatory relevance of the proposed mechanism—a point, they emphasized, that suggests that there is no reason to mix up the broader social mechanism approach with a narrower focus on "computer-based simulations".

In fact, analytical sociology's *manifestos* had already attempted to combine principles' elaboration with substantive research. In addition to Manzo's (2014) programmatic chapter discussed, and two other theoretical/methodological essays, Manzo's *Analytical Sociology: Actions and Networks* edited volume contained thirteen contributions that tried to answer well-defined empirical research questions. Among them, three studies relied on mathematical and/or agent-based computational simulations (without data), whereas ten others differently exploited historical data, survey data, geo-referenced information, digital traces, and/or lab or in-field experiments (see Manzo 2014: 12, 40–1). Edling and Rygren (2016: 1144, f. 1) surprisingly downplayed this fact and quickly claimed instead that "[…] although the volume edited by Manzo (2014) includes empirical chapters, they are almost exclusively preoccupied with simulations". Chattoe-Brown (2014) provided a more accurate analysis of these contributions. His conclusion was highly dismissive, however. In his eyes, with the exception of two studies coming very close to the entire sequence of steps defined by the seven postulates proposed by Manzo, the majority of contributions only focused on "some of these (or less charitably ignore others) with the result that they do not look much different in practice from existing forms of research". Manzo (2014: 10, 40) himself framed each chapter as a variation on the common theme represented by the seven proposed postulates defining analytical sociology's core: the substantive contributions were seen as providing theoretical and/or methodological resources and solutions to develop further and to implement specific elements of the research program. Chattoe-Brown did not accept this approach, and went straightforward to his ultimate goal: similarly, to the discontents with analytical sociology's *manifestos* I have evoked above, the reviewer wanted to deny that analytical sociology constitutes a "coherent research

Gianluca Manzo - 9781789906851

programme" that can pretend to be "rigorous in a *distinctive* way (rather than just being more rigorous than average)".

At present, we have a new resource to assess the extent to which substantive works exist that follow the principles and the methods progressively codified by analytical sociology's *manifestos*. In 2012, the *International Network of Analytical Sociology* decided to create the "Robert K. Merton Award" for the best paper in analytical sociology.[4] As the prize is awarded once a year at the annual meeting of the research network, eight articles have been awarded so far; in addition, three articles received an honorable mention in 2013, 2015, and 2020, respectively (see Table I.1). Compared to studies that are included within a special issue or an edited volume on analytical sociology, the interest of these eleven articles to evaluate analytical sociology at work is that they are not written to meet a specific invitation related to the endeavor of building analytical sociology. They were submitted to regular journals, and then submitted to the prize committee by their author(s) and/or nominated by third-party colleagues; they were finally evaluated by changing, and composite, prize committee members (see Table I.1). For this reason, in my view, the Merton award winning articles are an interesting source to assess whether or not analytical sociology has a core. Let me elaborate a bit more on this point.

As the Merton prize is supposed to recognize the "best" illustration of analytical sociology, authors must have some ideas on what analytical sociology is in order to decide to compete for the prize. They may be wrong, but it seems fair to suppose that, when they submit their paper, they self-assess their work as potentially relevant for the analytical sociology program. The members of the prize committees are in a similar position. They may have different perceptions of what analytical sociology is or should be, but it seems fair to assume again that they are guided by these perceptions when determining whether a paper is a good illustration of analytical sociology. Thus, the interest of these papers is that they allow us to see what analytical sociology is at the intersection of submitters' and evaluators' subjective understanding of analytical sociology. Should we find that at the intersection of these two sets of subjective perceptions the Merton award winning papers stably follow a similar set of principles, we may conclude that analytical sociology, at least within the network of prize candidates and evaluators, has a core. Should this core be in line with the principles discussed by analytical sociology's *manifestos*, we may further conclude that analytical sociology practices what it preaches.

Thus, in what follows, I will use this set of papers inductively and, by using as a benchmark the seven postulates I employed above to express synthetically analytical sociology's principles and methods, I assess how these principles and methods materialized in substantive works that their authors and various prize committees regarded as best illustrations of analytical sociology. In particular, after describing some formal signs of the article's academic visibility (section 2), I investigate the following aspects: (1) the presence in the Merton awarded articles

[4] The decision of naming the prize in honor of Robert King Merton wanted to signal the role played by Merton in the development of analytical sociology, both in terms of theorizing style and iconic models of social mechanisms (for a synthesis, see Hedström and Udehn 2009). A simple setup was designed for the prize. Both self-submission and third-party nomination were accepted; articles must have been published within the two academic years preceding the prize year; jury members were volunteers chosen among participants to the INAS annual meetings (or suggested by them); in 2016, the rule was introduced that the prize winner was an *ex-officio* member of the next year's prize committee; the prize committee had to work independently from INAS's council; honorable mentions were possible; the prize's call for submissions is usually advertised at the same time as the INAS annual conference's call for abstracts.

Gianluca Manzo - 9781789906851

Table I.1 Published journal articles having received the Merton award (or an honorable mention for it) by the International Network of Analytical Sociology between 2013 and 2020. For each paper, the following elements are shown: Prize committee composition, current number of citations, other academic prizes received, number of occurrences of the terms "analytical sociology" and "mechanism", and references to analytical sociology's manifestos

Prize Year	Winners	Citations	Other Prizes	#AS	AS *Manifestos*	#Mechanism
Prize Committee: Amir Goldberg, Andreas Flache, Daniel Alexandrov, Marc Keuschnigg (chair)						
2020 (a)	van de Rijt (2019)	8	–	0	–	6
2020 (m)	Leszczensky and Pink (2019)	16	–	0	–	39
Prize Committee: Ozan Aksoy, Arnout van de Rijt (Chair), Elizabeth Roberto, Thomas Grund						
2019	Goldberg and Stein (2018)	39	–	0	–	19
Prize Committee: Gunn Birkelund, Marc Keuschnigg, Gianluca Manzo (chair), Christian Stieglich						
2018	Aksoy and Gambetta (2016)	16	–	0	–	0
Prize Committee: Sarah Cowan, Delia Baldassarri, Rense Corten (chair)						
2017	Keuschnigg and Wolbring (2015)	55	1[5]	0	–	12
Prize Committee: Emily Erikson, Tim Müller, Christoph Stadtfeld (chair)						
2016	Cowen (2014)	59	1[4]	0	–	5
Prize Committee: Amir Goldberg, Clemens Kroneberg (chair), Paolo Parigi						
2015 (a)	Bruch (2014)	57	1[3]	0	–	10
2015 (m)	Skvoretz (2013)	47	–	1	H&S (1996, 1998); H(2005)	88
Prize Committee: Guillermina Jasso, Thomas Grund (chair), Arnout van de Rijt						
2014	Polavieja (2012)	41	1[2]	0	–	6
Prize Committee: Elizabeth Bruch, Christopher Edling, Henning Hillmann (chair)						
2013 (a)	DiMaggio and Garip (2011)	124		0	H(2005)	30
2013 (m)	Kroneberg and Wimmer (2012)	55	3[1]	1	H&B(2009)	19

Notes:
a=prize winner; m=honorable mention
[1] European Academy of Sociology Publication Prize (2013); ASA Rationality and Society Section Best Article Award (2013); Model Building and Simulation section Anatol Rapoport Award (2012).
[2] European Academy of Sociology Publication Prize (2013).
[3] Mathematical Sociology Section Outstanding Article Award (2015).
[4] Communication, Information Technologies and Media Section Best Paper (honorable mention) (2016).
[5] Model Building and Simulation section Anatol Rapoport Award (2014).

Gianluca Manzo - 9781789906851

of explicit references to analytical sociology and to the concept of mechanism (section 3); (2) the *explananda* of the Merton awarded articles and their scientific goal (section 4); (3) the way the Merton awarded articles design their explanatory machinery (section 5); (4) the way the Merton awarded articles combined data and methods to prove the empirical plausibility of the proposed mechanisms (section 6); (5) the presence in the Merton awarded articles' explanatory machinery of rational choice theories (section 7.1), and/or of "anomalous" theoretical elements (section 7.2).

2. MERTON AWARDS' SIGNS OF VISIBILITY

Before looking into the Merton awarded articles' content, let me first describe three formal indicators of the articles' academic visibility; namely, where they were published, the amount of citations they obtained, and their capacity to get other academic prizes.

As to publication outlets, one remarks that all the articles were published in highly visible and selective US-based or European journals. In particular, the eight prize winners appeared in *American Journal of Sociology* (4), *American Sociological Review* (1), *Sociological Science* (1), *Rationality and Society* (1), and the *European Sociological Review* (1). The three honorable mentions follow the same pattern: they were published in *American Journal of Sociology* (2) and in *American Sociological Review* (1).

As to citations, all the Merton awarded articles have already received some visibility on the citation market (see Table I.1). On average, until 30 November 2020, they received 47 citations (Google counts). With 124 citations, DiMaggio and Garip (2011), the "oldest" winner, is an outlier; with less than 20 citations, Aksoy and Gambetta (2016) and the two most recently awarded papers, both published in 2019, are the least quoted; the seven other papers are close to the average number of citations. Again, honorable mentions do not differ systematically from their winner companions—while Kroneberg and Wimmer (2012) is less quoted than DiMaggio and Garip (2011), Leszczensky and Pink (2019) is more quoted than van de Rijt (2019), and Skvoretz (2013) is relatively close to Bruch (2014). These figures should not be used to draw inferences on the relative quality, or impact, of each single paper. They are simply intended to substantiate the general observation that none of the Merton awarded articles went entirely unnoticed compared to the vast majority of journal articles that are published each year without ever being quoted, or only getting a few citations (for this regularity, across fields and times, see for instance, Van Noorden *et al*., 2014: 551; Kim *et al*., 2020: Table 1).

Finally, as to recognition from outside the *International Network of Analytical Sociology*, Table I.1 shows that five out of the eleven Merton awarded articles were awarded best article prizes by national professional associations (namely, the *American Sociological Association* and the *German Sociological Association*) and international science academies (namely, the *European Academy of Sociology*). Again, when comparing Merton honorable mentions with Merton prize winners, no systematic variation seems to exist: in one case, none of the two papers received another prize; in one case, the winner did, but not an honorable mention; and, in a third case, the symmetric situation occurred.

Thus, overall, the works finally recognized by the Merton prize committees as illustrating analytical sociology at its best cumulated a certain number of formal signs of academic visibility: they were all published in highly visible, mainly generalist, academic journals; they had already got some visibility on the citation market; and, for about half of them, they

Gianluca Manzo - 9781789906851

were also prized by academic instances other than the *International Network of Analytical Sociology*.[5]

3. MERTON AWARDS' FRAMING CHOICES

To move progressively from external signs of academic visibility to the articles' content, a preliminarily important question that should be addressed concerns the extent to which the Merton awarded articles explicitly embedded their substantive investigations within the meta-theoretical frames of analytical sociology and mechanism-based explanations.

Let me first consider the programmatic reference to "analytical sociology". Table I.1 (column "#AS") shows the number of occurrences of this expression in each article's abstract, main text, and footnotes. Surprisingly enough, it appears that the label "analytical sociology" is virtually absent from the Merton awarded articles. Only two of them employed it explicitly. In particular, Kroneberg and Wimmer (2012: 181) referred to analytical sociology when they explicated how their account of the formation of political alliances differed from existing macrostructural approaches, and claimed: "By contrast, our formal model follows the program of an analytical sociology (Hedström and Bearman 2009) and fully specifies all key mechanisms and assumptions". As to Skvoretz (2013: 486), the term appeared only generically in the article's abstract, where it stated that "In the spirit of analytical sociology, the author explores two mechanisms that could drive intra- and intergroup relations: attraction to similar versus repulsion from dissimilar others."

Table I.1 (column "AS manifestos") also shows that the virtual absence of the label "analytical sociology" from the Merton awarded articles translated into the scarcity of explicit references to the articles and books that attempted to progressively codify analytical sociology's principles and methods (see sections 1.1 and 1.3 above). Apart from Kroneberg and Wimmer (2012: 181) who explicitly quote Hedström and Bearman (2009) once, only DiMaggio and Garip (2011: 1888) quoted Hedström (2005) once; and Skvoretz (2013: 486) quoted Hedström and Swedberg (1998) and Hedström (2005) once. Leszczensky and Pink (2019: 396–7) may be considered a limit case in that they framed their use of stochastic actor-oriented modeling for longitudinal network data as "adding to the larger agenda of how social interactions among individuals result in aggregate outcomes" and quoted, for this *agenda*, Hedström and Manzo (2015)'s note on agent-based modeling as well as Keuschnigg et al.'s (2018) analysis of the relationships between analytical sociology and computational social sciences. References to substantive works from authors of analytical sociology's *manifestos* are not frequent either. These are limited to Hedström's earlier papers on spatial diffusion (see van de Rijt 2019: 1487), Hedström's chapter on rational imitation (see Polavieja 2012: 603; van de Rijt 2019: 1483, 1486), and Manzo and Baldassarri's (2015) study of the emergence of status hierarchies (see van de Rijt 2019: 1469–70).

[5] Let me emphasize that I am considering these three features from a purely descriptive point of view. My question here is only: do the articles that finally won the Merton prize possess some sign of academic visibility? From the answer to this question, which happened to be positive, I do not want to infer any causal effect on whether or not following analytical sociology's principles increases the chance of getting published in top journals, of being quoted, and/or of receiving a variety of academic prizes.

Gianluca Manzo - 9781789906851

However, it should be noted that, if explicit references to the analytical sociology's *manifestos* I discussed in section 1 are virtually absent from the Merton awarded articles, all of them contain at least one reference to concepts, models, and/or specific mechanisms that those *manifestos* have recurrently considered as part of the fundamental toolbox of analytical sociology. In particular, these are: (a) Schelling's iconic residential segregation model (see Bruch 2014: 1227; DiMaggio and Garip 2011: 1894) and his book *Micromotives and Macrobehaviours* (see Bruch 2014: 1227); (b) Coleman's contribution to mathematical sociology (see Skvoretz 2013: 492), analysis of the micro/macro loop (see Polavieja 2012: 602, 622), conceptualization of social capital (see Keuschnigg and Wolbring 2015: 102), as well as his creative analyses of diffusion through networks (see DiMaggio and Garip 2011: 1894; Cowen 2014: 468); (c) Merton's analysis of homophily (see DiMaggio and Garip 2011: 1892; Leszczensky and Pink 2019: 397) and of cumulative advantage mechanisms (see van de Rijt 2019: 1469); (d) White's model of vacancy chains (DiMaggio and Garip 2011: 1889); (e) Granovetter's theory of weak ties (see DiMaggio and Garip 2011: 1894; Keuschnigg and Wolbring 2015: 102) and threshold models (see DiMaggio and Garip 2011: 1894; Goldberg and Stein 2018: 900; van de Rijt 2019: 1469); (f) Boudon's analysis of educational inequalities (DiMaggio and Garip 2011: 1889), of generative *versus* statistical models (see Skvoretz 2013: 488), and of rational action (see Polavieja 2012: 622); (g) Elster's studies of norms (see Polavieja 2012: 602) and rational self-commitment (Aksoy and Gambetta 2016: 793) as well as Elster's treatise on *Nuts and Bolts* (see van de Rijt 2019: 1470). The recurrence of this set of analytical sociology's relevant references throughout the Merton awarded articles clearly conveys a sense of family resemblance across the articles.

Let me now consider the use of the basic meta-theoretical building-block of analytical sociology, i.e. the concept of mechanism. As shown in Table I.1 (column "#Mechanism"), with one exception, all Merton awarded articles explicitly and repeatedly employed the term "mechanism" to refer to their explanatory machinery. This is indeed the linguistic element that really generates a sense of unity across the Merton awarded articles. However, variation around the average number of occurrences (21) is substantial, and marked differences in the level of explicitness of the definition of the concept of mechanism appear across the articles.

In particular, on the one hand, the reader finds DiMaggio and Garip (2011: 1888), who explicitly endorsed a definition of a "social mechanism" as entailing (a) goal-directed individual actions and (b) consequent social interactions that (c) yield higher-level outcomes that (i) are emergent (i.e. that cannot be recovered simply by aggregating individual actions that combine to produce them) and (ii) vary depending on the initial social structure (ordinarily depicted in terms of social networks). For this understanding of a mechanism, they referred to Hedström (2005), and, interestingly, to Tilly's book on *Identities, Boundaries, and Social Ties*. Similarly, Skvoretz (2013: 486) explicitly quoted the definition proposed by Hedström (2005: 11), according to which a social mechanism can be generically understood as "a constellation of entities and activities that are organized such that they regularly bring about a particular type of outcome". On the other hand, the reader finds Aksoy and Gambetta (2016) who did not even use the term "mechanism," even though, as we will see in section 5, the theoretical hypotheses they wanted to test empirically were explicitly, and transparently, formulated in terms of actors' preferences, their potential strategic choices, and their reactions to proximate (namely, the neighborhood an actor lives) and distant (namely, norm and value macro-changes) social contexts. The remaining Merton awarded articles, hence the majority of them, fall between these two extreme cases. They recurrently employed the term "mechanism" without

Gianluca Manzo - 9781789906851

explicitly defining it. However, as I will show in section 5, this did not translate into a rhetoric use of the concept of mechanism. Each article in fact described a well-described chain of events unfolding at different scales, these events being fuelled by the actions and interactions of equally well-specified entities who follow clearly identified logic, and this under overtly explicated scope conditions (for a summary, see Table I.2's columns "Explanans"). In this sense, no matter whether explicitly defining the term or not, all the Merton awarded articles employed the concept of mechanism in agreement with analytical sociology's manifestos, which progressively converged toward generically seeing a mechanism as a set of organized entities whose properties and activities are able to trigger changes that generate the observed connections with some regularity (see Hedström 2005: 25; Manzo 2010: 138; Manzo 2014: 14; Keuschnigg et al. 2018: 5).

Thus, overall, the works finally considered by the Merton prize committees as illustrating analytical sociology at its best shared a paradox: the majority did not explicitly refer to analytical sociology's *manifestos* to frame their substantive investigations and, while they employed the term "mechanism" widely, they did not define the term explicitly. It is the use of secondary references (i.e. references to concepts, models, and typical mechanisms that analytical sociology's *manifestos* recurrently regarded as founding pieces of the analytical sociology's toolbox) as well as the way concrete mechanisms are designed that gave the Merton awarded articles a sense of unity and of family resemblance. One must, then, carefully inspect the inner working mode of these articles to assess the extent to which they illustrated the principles and the methods of analytical sociology.[6]

This is what I do in the next four sections, where I inspect Merton awarded articles' explananda (section 4), the structure and the content of their *explanans* (section 5), the way the *explanans* is connected to the *explanandum* through data and methods (section 6), and what type of action and structural theories are mobilized (section 7).

4. MERTON AWARDS' *EXPLANANDA*

Analytical sociology's *manifestos* have repeatedly presented analytical sociology as primarily concerned with the explanation of puzzling "collective properties" (see Hedström 2005: 67) or "social facts" (Hedström and Bearman 2009: 3). As we have seen in section 1, the first two postulates of the analytical sociology's core research strategy as synthetically defined by Manzo (2014: 7) are all about the preliminary conceptual and the empirical identification of these facts. Manzo (2014: 9, 38) even characterized the overall intellectual project of analytical sociology as "an empirically oriented, experimentally and computationally based, *macro-sociology* with clearly explicated and empirically grounded dynamic micro- and

[6] It seems very difficult to ascertain empirically the mix of motives behind the choice of not referring explicitly and extensively to analytical sociology's *manifestos*. It may be that authors want to save space for substantive analyses and methodological developments rather than discussing research principles on a meta-theoretical level; it may also be that authors expect reviewers of certain journals to dislike this type of programmatic discussion. Authors may also act strategically. Depending on how they perceive the academic reputation of analytical sociology, they may consider that explicitly mentioning analytical sociology may reduce, rather than increase, the chance their substantive analyses are objectively assessed; they may also associate analytical sociology with specific authors to whom they may not want to give credit. I will not pursue these speculations here. The important point is to highlight the paradox. As I will argue in the concluding section, being aware of it is the first step to appreciate its potential negative consequences.

Gianluca Manzo - 9781789906851

Table I.2 Merton-awarded articles' explananda and explanans

Winners	Field	*Explananda*	*Explanans*		
			1. Context	*2. Action*	*3. Transformation*
van de Rijt (2019)	Sociology of quality, reputation, and status construction	Temporal persistence of status hierarchies with quality/status mismatch	Various options with uncertain quality	Others' choices help to reduce uncertainty (rational imitation)	Herding behavior
Leszczenskya and Pink (2019)	Sociology of networks and sociology of race and ethnicity	Ethnic homophily in school-based friendship networks	Ethnic group size and distribution of ethnic identification	Positive social identity maintenance	Ethnic identification; interdependence at dyadic level
Goldberg and Stein (2018)	Cultural and cognitive sociology	Patterns of clustered cultural preferences	Distributions of observable behaviors	Strive for cognitive consistency	Looking for practice co-occurrences in others' behavior
Aksoy and Gambetta (2016)	Sociology of religion	Social patterns of veiling in Muslim dominant and non-dominant contexts	Increasing women's education, employment, and urbanization	Signal to others one's moral integrity and piousness	Reactions to social contacts potentially threatening one's moral integrity
Keuschnigg and Wolbring (2015)	Sociology of deviance and social norms	Variation in norm compliance as a function of contextual cues and social capital	Community-level social capital; traces of past norm violations	Need to assess conviction probability; desire to escape sanctions	Prior norm violations change incentives to further norm violations
Cowen (2014)	Political sociology and social opinion changes	Patterns of moral attitudes toward value-sensitive practices and events	Distribution of miscarriage/abortion; distribution of political attitudes	Stigmatization avoidance tendency	Interdependence between disclosure choices and hearing attitudes
Bruch (2014)	Sociology of inequality	Changes in income inequality and residential segregation by race	Race group size, within- and between-race income inequality	Mobility choices depend on income and in-group preferences	Prior moves change neighborhoods' desirability

(continued)

Gianluca Manzo - 9781789906851

Table I.2 (continued)

Winners	Field	*Explananda*	*Explanans*		
			1. Context	*2. Action*	*3. Transformation*
Skvoretz (2013)	Sociology of networks and of family	Homophily patterns in marriage, dating, and cohabitation	Group size sets the encounter opportunity structure	Strive for similar others; probabilistic rejection dissimilar	Interdependence of attraction and rejection choices
Polavieja (2012)	Sociology of inequalities and job market stratification	Gender stratification in job-specific skills	Aggregate job rates, fertility rates and domestic division of labor	Others' experiences signal chances of success (Rational imitation)	Prior choices and outcomes within ego's proximate social environment
DiMaggio and Garip (2011)	Sociology of inequalities, diffusion, and migration	Patterns of intergroup inequalities in technology adoption and migration	Differential resources/ Homophilous ego's social network on resources	Following others' choices provide utility gains and information	Prior choices by members of ego's social network
Kroneberg and Wimmer (2012)	Historical and political sociology	Trends in the structure of political alliances	Differential resources/ Power asymmetry	Strive for resource control and cultural similarity	Dynamic sequence of offers and counter-offers

Gianluca Manzo - 9781789906851

network-level foundations". As a minimal, working definition of what "collective properties" or "social facts" are, analytical sociology progressively converged to consider "macro", what cannot be defined, measured, or understood with reference to a single member of the population of interest (Ylikoski 2012: sec. 2.3). For this reason, phenomena of interest of analytical sociology, and more generally of scholars interested in mechanism-based explanations, have also been generically defined as structural properties (see Marchionni and Ylikoski 2013: 334).

Table I.2 (column "*Explananda*") synthetically shows that each Merton awarded article focused on a well-identified "structural property" with the explicit goal of explaining its emergence and/or temporal stability/instability. These "structural properties" touched upon a large spectrum of macro phenomena, and spanned across a variety of specialized sociological fields (see Table I.2, column "Field").

In particular, DiMaggio and Garip (2011: 1899, 1916) contributed to sociology of inequalities, of the diffusion of innovations, and of migration by studying the persistency over time of intergroup inequality in at-home internet use in the US between 1997 and 2007 and, on the other hand, the variation over time of rates of intergroup rural-urban migrations among 22 rural villages in Thailand between 1972 and 2000. Kroneberg and Wimmer (2012: 177, 198, 220–1) wanted to explain macroscopic patterns of historical political changes "[…] from the estate model of society under the *ancien régime*, through populism, to fully inclusive nation building", with a particular focus on France (1300–1900) and the Ottoman Empire (1500–1900), thus contributing to comparative historical sociology of nation-state formation and political sociology. Polavieja (2012: 594, 620) aimed at understanding gender-based patterns of investments in job-specific skills in 234 different regions of 24 European different countries, thus contributing to the field of sociology of inequalities and job market stratification. Skvoretz (2013: 489) wanted to explain variations in levels of homophily by nationality, ethnicity, race, education and religion in marriage, dating, and cohabitation as observed in many societies, including the United States, Great Britain, Austria, and 21 other European nations, thus contributing to sociology of networks and couple formation.

Bruch (2014: 1223–4, 1246) contributed to stratification research by studying the temporal interdependence between the size of race groups, in particular white and blacks, the levels of income inequality within and between those groups, and the resulting patterns of racial residential segregation in major American cities, including Los Angeles, Atlanta, and Chicago, between 1980 and 2000. Cowen (2014: 466) contributed to the sociology of political opinion and social change by analyzing the connection between patterns of hearing and disclosure of private information on emotionally and value-loaded women's subjective life events, like miscarriage and abortion, and the persistence of liberal/conservative attitudes toward these events among contemporary American adults. Keuschnigg and Wolbring (2015: 97, 98, 103) wanted to understand spatial variations in the compliance to social rules for low-cost behaviors (like littering, jaywalking, or stealing small amounts of lost money) as a function of context-specific cues (of existing levels of norm violation) and varying amounts of neighborhood-level social capital, thus contributing to sociology of deviant behaviors and social norms. Aksoy and Gambetta (2016: 372) aimed at understanding social patterns of Muslim women's propensity to veil (in Belgium, Turkey, and 25 other Muslim countries) as a function of those women's socio-economics attributes, levels of religiosity, and the type of neighborhood they leave in, thus contributing to sociology of modernization and religion as well as to political sociology. Goldberg and Stein's (2018: 897, 903, 908) study fell within the field of cultural and cognitive sociology: they wanted to understand how individuals progressively come to associate

Gianluca Manzo - 9781789906851

specific subjective meanings to given cultural practices and how, once these meanings are in place, cultural practices are endorsed as clusters of practices rather than isolated practices, thus making cultural differentiation a cognitively embedded macro phenomenon. Leszczensky and Pink (2019: 394, 397) studied levels and temporal variations of segregation by ethnic groups in friendship networks at school among German young adolescents as a function of students' ethnic identification, thus contributing to sociology of social networks and ethnicity. Finally, van de Rijt (2019: 1469–70, 1472) contributed to sociology of quality, reputation, and status construction by explaining the potential temporal instability of status hierarchies where inferior options initially outperform superior ones as a function of the strength of the social influence affecting the individuals who have to choose these various options.

Thus, the Merton awarded articles well illustrated the *a*-specific nature of analytical sociology. Topics of interest of analytical sociology are not defined *a priori*. Analytical sociology is a generic sociological perspective: it is not tied to a specific substantive field. *Explananda* can be of any kind provided they concern "structural properties", i.e. cross-sectional and/or dynamic patterns of behaviors, practices, or outcomes across socio-economics/demographic groups, and/or social contexts that cannot be understood, measured, and explained by reference to any single member of the population under exam.

The Merton awarded articles nicely showed that these structural properties can be points of equilibrium of a social system—like cultural differentiation in Goldberg and Stein (2018)—transitions from one equilibrium to another—like historical changes in political regime modelled by Kroneberg and Wimmer (2012) or unstable status hierarchies analyzed by van de Rijt (2019)—or still dynamic flows—like migrations in DiMaggio and Garip (2011). Structural properties of interest can also be associations between contextual features at one point in time and distributions of attributes or choices at a subsequent point in time—like the dependence of gender-based investments in job-specific skills on gender-based occupation rates studied by Polavieja (2012), the impact of contextual cues and level of social capital on deviant behaviors analyzed by Keuschnigg and Wolbring (2015), or levels of modernization and variation in religious behaviors inspected by Aksoy and Gambetta (2016). Structural properties can also be associations between population distributions—like the association between income distribution and residential segregation studied by Bruch (2014) or the dependence of the distribution of political opinions and that of value-loaded practices and events analyzed by Cowen (2014). In some cases, structural properties of interest concerned the configuration of social ties, without further moving on to a higher level of abstraction—like the various forms of network homophily studied by Skvoretz (2013) and by Leszczensky and Pink (2019).

No matter the specific content and form of the structural properties of interest, all Merton awarded articles clearly formulated the *explanandum* in a transparent and pedagogical way, typically in the article's introductory section. The description of the *explanandum* is always associated to a *why*-question: the main goal is to understand the source of the structural properties of interest. Sometimes, the *why* question is formulated in terms of "puzzles", i.e. an intriguing fact that arises from the contradiction between theoretical expectations and empirical observations, or an unexpected contradiction between facts themselves (see, for instance, DiMaggio and Garip 2011: 1988, 1895, 1913, 1914; Kroneberg and Wimmer 2012: 179; Polavieja 2012: 592–3; Aksoy and Gambetta 2016: 792, 803; Goldberg and Stein 2018: 898). In all cases, the answer to the *why*-question takes the form of a *how*-solution: the bulk of each Merton awarded article consists in the formulation of a chain of events unfolding at different scales, these events being fueled by the actions and interactions of well-specified entities who

Gianluca Manzo - 9781789906851

follow clearly identified logics, i.e. a mechanism operating under overtly explicated specific conditions.

As we will see in the next section, while the specific content of this mechanism varies from one article to another, its structure is relatively stable. In particular, a common feature of each proposed mechanism is action interdependence. No Merton awarded article considered that the simple accumulation of individual actions unfolding independently from one another was sufficient to make the macroscopic outcome of interest transparent.

5. MERTON AWARDS' *EXPLANANS*

Analytical sociology's *manifestos* formulated various typologies of social mechanisms. In particular, Hedström and Swedberg (1996: 296–8; 1998: 22) initially followed the three main branches of the Coleman's diagram and proposed to distinguish between "situational", "action-formation", and "transformation" mechanisms. Hedström (2005: chapter 3) added specificity to this typology by formulating, on the one hand, a generic sub-typology of mechanisms involving dyadic interactions (ibid., 47–56) and, on the other hand, a list of typical action-formation mechanisms involving various patterns of opportunity-beliefs-desires connections (ibid., table 3.1).

The explicit use of these typologies of social mechanisms is virtually absent from the Merton awarded articles. Polavieja (2012: 603, 622) is the only exception: he directly relied on "rational imitation" (one of the action-formation mechanisms of Hedström's typology) and explicitly, though generically, referred to the concept of "situational mechanism". However, the explanations proposed by each Merton awarded article are fully in line with these typologies. Indeed these *explanans* systematically combined three aspects: (1) a description of the specific pathways through which the context in which actors shape their beliefs, preferences, and/or opportunities (i.e. situational mechanisms); (2) the principles underlying actors' actions (i.e. action-formation mechanisms); and (3) how action interdependency progressively generates the macro outcome of interest (i.e. transformation mechanisms). The three columns of Table I.2 under the common heading "*Explanans*" aim to help the reader synthetically appreciate these explanatory elements, thus facilitating cross-article comparisons. In what follows, I will try instead to restitute as neatly as I can the details of each proposed explanation. To do so, I will consistently use 1, 2, and 3 to refer to elements related to situational, action-formation, and transformation mechanisms, respectively. When more than one elements substantiate a given mechanism, I will adopt multilevel lists (for instance, 1.1, 1.2 (…), 2.1, 2.2, and so on).[7]

Let me start with Kroneberg and Wimmer (2012: 177, 187–9), who wanted to explain how political modernization can lead to a national community, to a set of segmented ethnic groups, or to populism. To do so, they framed the problem in terms of typical political alliances that are more likely to emerge. The proposed mechanism driving the formation of these alliances is organized as follows: (1) actors (collective actors, in this case) have different resources (as captured by taxing capabilities and provision of public goods), different power (as captured by military resources), different cultural identities (as captured by religion, language, skin

[7] I thank Viviana Amati for this suggestion and for noting symbol inconsistency in a previous version of the text.

Gianluca Manzo - 9781789906851

color, and various cultural dispositions) as well as different distance from the center of the political decision making; (2) giving this resource distribution and power asymmetry, to form their preferences for potential allies, the actors' goal is to find the best balance between control over resources and the ally's cultural proximity; (3) in this context and given these preferences, central elites make proposals, peripheral elites make counter-proposals, and masses decide who they want to support, thus creating a dynamic sequence of offers and counter-offers that will stabilize at some point when no actor can find a better ally. Kroneberg and Wimmer (2012: 177–8, 208, 215, 2017) clearly identified the conditions under which the posited "struggle" is more likely to end with a given alliance system—the stronger the state and the denser the network of civil society organizations, the higher the probability that inclusive and consensual alliances and identities will emerge.

DiMaggio and Garip (2011: 1888–9, 1893, 1900) wanted to explain the generation of inequalities in internet adoption among individuals with different incomes and education, as well as over-time divergent rural-urban migration rates between Thai villages. To do so, they proposed a common mechanism made of the following elements: (1.1) actors possess different amount of resources to sustain the costs of subscribing internet or to migrate; (1.2) actors experience proximate social networks that tend to be homophilous with respect to socio-economic factors; (2) when making up their mind about whether to adopt or migrate, actors look at prior adoption choices among their relevant direct contacts (i.e. friends or family members)—because having friends using the Internet increases the benefits of having it, and having migrant friends or family members provide information and logistic help to migrate (ibid., table 1); (3) relying on choices of proximate others with similar amount of resources will amplify non-adoption among have-nots and adoption among haves. DiMaggio and Garip (2011: 1888, 1891) clearly stated the condition under which the posited mechanism is likely to generate intergroup inequality: network homophily must exist on factors that are associated with adoption.

To explain the underrepresentation of women in highly specialized jobs, Polavieja (2012: 592–3, 600, 603) proposed the following mechanism: (1.1) actors' proximate social environment is rich in cues informing them about their job survival risks in different investment paths; (1.2) in particular, actors can look into previous generation's outcomes, domestic labor division in peers' couples, and peers' fertility behaviors; (2) actors rely on others' past and contemporary experiences to roughly estimate their survival probability in a given job; (3) by looking at others' past and contemporary experiences, actors who are surrounded by proximate social environment with less favorable signals about their chances of success in jobs requiring specific skills will be discouraged to follow this paths on the job market. Polavieja (2012: 594) posited that this mechanism is more like to operate for groups like women that face greater uncertainty and higher risks when assessing different job market options.

Skvoretz (2013: 491–3, 499) wanted to explain the variations of homophily on various attributes in different types of couple formation relationships and, to this aim, he proposed a mechanism involving the following elements: (1) the demographic size of groups provide actors with certain opportunities for encountering similar and dissimilar others; (2) actors' are driven by a tendency toward homophily so that their preference for similar others bias the structural constraints; in particular (2.1) when someone meets a similar individual, then the tie is always formed (attraction mechanism); otherwise, (2.2) when someone meets a dissimilar other, the tie is rejected with a given probability (repulsion mechanism); (3) one's attraction/rejection choices dynamically change the opportunity to encounter structures for

Gianluca Manzo - 9781789906851

subsequent attraction/rejection choices. Depending on the initial group size, and the intensity of the rejection bias, variations in the level of homophily along different attributes emerge in different settings. Skvoretz (2013: 500) identified the following scope condition: "repulsion is the driver of intergroup relations when the dimension type is ascribed, like ethnicity/race and religion, while attraction is the driver of intergroup relations when the dimension type is achieved, like educational attainment".

To explain the complex three-way interdependence between the distributions of income and race, and patterns of racial residential segregation, Bruch (2014: 1221–2, 1243–5) built the following mechanism: (1.1) actors belong to race groups with different sizes, possibly changing over time; (1.2) they also have income that is variable both between race groups and within them, and possibly changing over time; (2) actors move from one neighborhood to another as a function (among other factors) of their income and the race composition of the potential new destination, with relatively mild in-group preferences; (3) each family's residential move changes the race composition of neighborhood for subsequent choices. When income and race are correlated, the dynamic interplay of sequential moves generates offsetting effects where affluent blacks may experience larger exposure to affluent white neighbors whereas poor blacks may find themselves even more intensely isolated within low-income black neighborhoods. Bruch (2014: 1224, 1263, 1264) identified the conditions under which offsetting is more likely to happen—namely sufficiently high within-race income inequality, increasing between-race income inequality, and sufficiently high proportion of blacks—and clarified that "whether these offsetting processes result in a net increase or decrease in segregation depends on the relative size of the black population, the salience of racial versus economic factors in residential mobility decisions, and the shape of the income distribution" (ibid., 1224).

To explain how patterns of secret hearing and disclosure impact population-level opinion changes, Cowen (2014: 471–2) described the following mechanism: (1.1) at any given time, a stock of private events (miscarriage and abortion, here) exists in the population; (1.2) women personally experiencing these events may decide to share their experience to relevant confidants like friends, family members, colleagues, or more distant acquaintances; (2) when disclosing secrets, actors want to avoid stigmatization; (3) as a consequence, private events that are less likely to be disapproved (miscarriage rather than abortion) will be more frequently disclosed, and they will be more frequently disclosed to those confidants that are perceived as having more favorable views on the to-be-disclosed event (namely, those with liberal rather than conservative political attitudes). The dynamic interdependence between disclosure choices and hearing attitudes is likely to create what Cowen (2014: 495) calls a "self-fulfilling illusion"—more stigmatized events will be less present in actors' subjective perceptions than they actually are in the population, and those who are more likely to disapprove the events will be less likely to be aware of them—which is likely to slow down attitude changes. As a by-product, Cowen's (2014: 484) mechanism identified awareness of others' behavior as a necessary condition for social influence: "Selective disclosure and secret keeping […] are mechanisms for thwarting social influence by preventing the awareness that is a precondition of influence […]".

To explain variations in levels of norm compliance across local settings, Keuschnigg and Wolbring (2015) designed the following mechanism: (1.1) actors live in local context with varying levels of others' supportiveness, quality of social relationships, and shared activities

Gianluca Manzo - 9781789906851

(i.e. "social capital"); (1.2) social environment also contains varying amount of traces (like physical or social disorder) of previous norm violations; (2) when making up their mind about whether to comply with a norm, actors exploit contextual cues as an heuristic to infer the probability of being sanctioned; (3) looking at previous violations of a given norms (litter, for instance) can lead actors to further violation of the same norm ("same-norm inhibition effect"), and also to feel more comfortable with violating other norms ("cross-norm inhibition effect"); for instance, by stealing a small amount of in-the-street lost money after seeing littering. Depending on the initial stock of violations displayed in a given local context and the amount of social capital, temporal interdependence between norm violation choices will lead to varying levels of norm compliance. Keuschnigg and Wolbring (2015: 99) clearly formulated a scope condition for the mechanism: the smaller the norm to be violated (i.e. a norm associated with minor sanctions), the more the actors rely on contextual signals rather than on lengthy cost-benefit evaluations. Reasoning that past norm violations within a civic environment are likely to be more informative than within those that are structurally disordered, Keuschnigg and Wolbring (2015: 103) also identified a moderating condition for the mechanism: the effects of contextual signaling are likely to be stronger when social capital is high.

To explain variations of veiling behaviors among Muslim women in Muslim dominant and non-dominant contexts and, in particular, the possible counter-intuitive positive correlation between modernization and veiling frequency, Aksoy and Gambetta (2016: 793, 796) posited the following mechanism: (1) modernization leads Muslim women to experience more education, greater access to job market, and to urban settings; (2.1) Muslim women may veil because they believe this is a religious requirement; but (2.2) they may also veil to signal their moral integrity to members of their home community as well as to new social contacts outside this community. If so, (3.1) modernization, by making religious beliefs less prominent, may reduce veiling's average frequency while, (3.2) by exposing Muslim women to more secularized settings may increase the propensity of veiling among women who remain highly religious, and who are highly educated, urban, and middle class; (3.3) highly religious women living in urban neighborhoods where veiling is less frequent and more various in terms of dressing manifestations, as well as highly religious women who live in neighborhoods where Muslims are a minority, should also manifest a higher propensity to veil because these two types of neighbourhoods are likely to increase the necessity for women to signal their moral integrity and piousness.

To explain the emergence of clusters of cultural practices without assuming social network segmentation, Goldberg and Stein (2018: 908) developed the following mechanism: (1) actors have an initial set of perceptions about how strongly practices are associated to one another; (2.1) actors have initial preferences for a set of practices; (2.2) actors endorse and exhibit a given practice proportionally to the strength of the preference for that practice; (2.3) as co-occurrences of practices are observed, actors update their preferences for given practices as long as the practices at hand are compatible with the current state of their set of perceived associations among practices; (3) actors observe other actors' practices and, when they observe that two practices are endorsed by a given actor, they reinforce their own perceived association between the observed practices. Goldberg and Stein (2018: 906–7) called this mechanism "associative diffusion" and demonstrated that it is sufficient to generate agreement among agents on which practices go together while allowing agents to have different preferences for specific practices (ibid., 914). They also formulated the condition under which "associative

Gianluca Manzo - 9781789906851

diffusion" is more likely to apply, namely when behaviors are observable and when the utility of adopting a given practice is uncertain (ibid., 921–2).

To explain same-ethnic friendship, and thus ethnic segregation in friendship networks at school, Leszczensky and Pink (2019: 396–7, 399–401, 403) combined the following elements: (1) the size of ethnic groups structurally constrain the opportunity to form ties with students from a given ethnic background; (2) given these constraints, students' are sensitive to potential friends' ethnicity but they also try to preserve a positive social identity by not befriending someone with markedly different levels of ethnic identification; (3.1) when this goal is in place, high identifiers may prefer same-ethnic friends provided they are also high identifiers, while excluding low identifiers even though they may come from the same ethnic group; (3.2) symmetrically, low identifiers may avoid to befriend high identifiers irrespectively of their ethnicity; (3.3) preferential attachment, exclusion, and avoidance are expected to operate under dynamically evolving levels of ethnic identification, because adolescents may adjust their own level of ethnic identification to that of their friends.

Finally, to explain the potential instability of status hierarchies where low-quality options are more popular than high-quality ones, van de Rijt (2019: 1469–70, 1472) proposed the following mechanism: (1) in the actors' proximate social contexts, several options (goods, ideas, opinions, etc.) exist, and (1.2) some actors have already chosen one of these options; (2) to assess the option's quality, actors' are positively and monotonically (but probabilistically) sensitive to the proportion of people choosing a given option so that the higher that proportion, the stronger the preference for the option at hand; (3) under these conditions, the smaller the initial proportion of those choosing inferior options, the higher the chance that some individual deviates from this choice, thus progressively triggering a new majority leading to the correct evaluation of superior alternative. Van de Rijt (2019: 1470, 1486) identified the "magnitude of the social influence effect" as the main condition under which status can be decoupled by quality: unless this effect is unrealistically high, he claimed, a self-correcting dynamic will always have a sizeable chance to happen.

Thus, behind the variety of the specific content of each *explanans*, all Merton awarded articles followed a common logic to design this *explanans*: (1) they described how contexts shape actions; (2) they explained why actions are performed; and, (3) they described how action interdependence led to high-level patterns. In this sense, the Merton awarded articles all instantiated the third and fourth postulates of the analytical sociology's core research strategy as synthetically defined by Manzo (2014: 9) (see section 1.3). Moreover, in spite of the absence most of the time of an explicit definition of the concept of mechanism (as we have seen in section 3), and, as I have noted above, without directly referring to typologies of mechanisms proposed by analytical sociology's *manifestos*, the Merton awarded articles all provided a "mechanism" that fully meets the generic definition of a mechanism toward which those *manifestos* progressively agreed over the years, which see a mechanism as a set of organized entities whose properties and activities are able to trigger changes that generate the observed connections with some regularity (see Hedström 2005: 25; Manzo 2010: 138; Manzo 2014: 14; Keuschnigg et al. 2018: 5). Of particular interest is the fact that the conditions under which the proposed mechanism is expected to operate are always clearly formulated. This was one of the critiques addressed to early analytical sociology's *manifestos*: the lack of attention for scope and triggering conditions (see Gross 2009: 373; for a reply, see also Manzo 2010). The Merton awarded articles avoided this shortcoming.

Gianluca Manzo - 9781789906851

Table I.3 Merton-awarded articles' methods, data, and selected aspects of the theoretical framework

Winners	Methods	Data	Rational action	Anomalous concepts
van de Rijt (2019)	Choice formation mathematical model and large-scale, web-based experiments	Experimental data	Others' choices convey information on quality	–
Leszczensky and Pink (2019)	Stochastic actor-oriented models for network formation and behavior (SAOM)	Longitudinal ego-centered network data	–	Tajfel's theory of social identity
Goldberg and Stein (2018)	Agent-based modeling	–	Others' choices convey meanings on practices	Meaning-centered perspective
Aksoy and Gambetta (2016)	Structural equation modeling	National and comparative cross-sectional survey data	Signaling and commitment strategies	–
Keuschnigg and Wolbring (2015)	Field experiments and questionnaires (for social capital measurement)	Experimental data and survey data	Others' choices convey information on risk of conviction	–
Cowen (2014)	Tailored, internet-based survey and logistic regressions	Cross-section survey data and qualitative data from health system	–	Meaning-centered perspective
Bruch (2014)	Empirically calibrated agent-based modeling and various multivariate statistical models	Longitudinal survey data and Census data	–	Blau's social structure theory
Skvoretz (2013)	Tie formation mathematical model + statistical estimation of mechanism-related parameters	Cross-section comparative survey data	–	Blau's social structure theory

(continued)

Gianluca Manzo - 9781789906851

Table I.3 (continued)

Winners	Methods	Data	Rational action	Anomalous concepts
Polavieja (2012)	Investment decision mathematical model and hierarchical regression models	Cross-section, comparative survey data	Others' choices convey information on survival probability	–
DiMaggio and Garip (2011)	Empirically calibrated agent-based modeling and various multi-variate statistical models	Cross-section and longitudinal survey data	Others' choices convey information on attractiveness	–
Kroneberg and Wimmer (2012)	Empirically-calibrated exchange-theoretic model + sequential, non-cooperative game-theoretic model	Archive historical data	Weak rationality, endogenous preferences and cultural motives	Tajfel's theory of social identity

Note: "–" means that the element at hand is absent from the paper.

Gianluca Manzo - 9781789906851

6. MERTON AWARDS' DATA AND METHODS

How did the Merton awarded articles combine data and methods to persuade the reader that the *explanans*, i.e. the postulated mechanisms that I reconstructed in the previous section, can be regarded as realistically at work in the real world?

In section 1 we saw that analytical sociology's *manifestos* set a progressively more precise and demanding methodological *agenda* for mechanism-oriented social research. The field started from a critique of the explanatory deficiencies of correlation-oriented multivariate statistics (see Hedström and Swedberg 1998: 15–17; Hedström 2005: chapter 5), moved on to an increasing emphasis on combining statistical methods with agent-based computational simulations (Hedström 2005: Ch. 6; Manzo 2007; Hedström and Bearman 2009: 16), and, as implied by Manzo's (2014: 9) postulates 5–7 (see section 1.3) ended up with proposing "a complex interface among multivariate statistics, computational methods, mathematics, and experiments in which each method is mobilized to accomplish specific tasks" (Manzo 2014: 37). Moreover, the complementarity of methods pushed analytical sociology's methodological *agenda* to value the importance of combining different types of qualitative, quantitative, experimental, and small and big data with the ultimate goal of finding the best way to document the empirical traces of the postulated mechanisms (see Manzo 2014: 12, 37, 40; Manzo et al. 2018: 1146; Keuschnigg et al. 2018).

How do the substantive works that were awarded the Merton prize for illustrating analytical sociology at its best fit this picture? The second and third columns of Table I.3 provide a synthetic answer to this question, thus facilitating cross-article comparisons. In what follows I consider instead the methodology of each Merton awarded article in detail, again from the oldest to the most recent prize.

To study how different systems of political alliances emerge, Kroneberg and Wimmer (2012: 180, 187–9, 191, 193, 196) relied on a mathematical model that combines, on the one hand, an exchange model formalizing how actors form their preferences over possible alliance configurations and, on the other hand, a sequential, non-cooperative game model formalizing how negotiation between actors unfold. Crucial parameters of the exchange theoretic models, namely resource control and power, are based on "carefully researched historical data on the distribution of taxing capabilities, public goods provision, and military support in France (1300–1900) and the Ottoman Empire (1500–1900)" (ibid., 181–2), which, they claimed "prevent us from simply assuming the parameter values that will generate the hypothesized outcomes". Given this empirical calibration, Kroneberg and Wimmer (2012: 218) observed that the model was able to generate the rhythm and the outcome of political identities and alliances historically observed in France and the Ottoman Empire, namely more inclusive national building and less ethnic enclosures and status differentiation in the former case. That said, being aware that full empirical calibration and outcome validation was not possible for the long-term historical process of interest, Kroneberg and Wimmer (2012: 182–3, 201–2, 214, 217, and Appendix C) also performed extensive sensitivity analysis (on the empirical values adopted for the exchange model's parameters) and various robustness checks (on various model's assumptions, like the number of actors, the sequences of moves, political decisions, or the distributions of cultural traits), showing that results were qualitatively unchanged.

To demonstrate that the dependence of current choices on prior choices within homophilous social networks can lead to intergroup inequalities in internet adoption as well as to inter-village differences in Thai rural-urban migration rates, DiMaggio and Garip (2011) adopted

Gianluca Manzo - 9781789906851

two different methods. In the former case, they designed an agent-based computational model depicting how *ego*'s reservation price to adopt the Internet is impacted by others' choices (ibid., 1894). Crucial agent-level parameters (like the number of social contacts, income, race, and education) are based on US nationally representative survey data (ibid., 1900–1901). Given the inequality patterns by race and education observed in survey data between 1997 and 2007 (ibid., 1899), the model is simulated under different types of network externalities and varying levels of network homophily to establish which conditions led to aggregate outcomes that qualitatively resembled the observed ones (ibid., 1902 and ss.). Multivariate statistical models were estimated to describe the simulation results both at the agent- and system-level (ibid., 1908–1911). In the Thai case, DiMaggio and Garip (2011: 1916, 1920–1) attempted to find direct traces of the inequality effect of network homophily by exploiting detailed longitudinal survey data describing migration choices and flows from 22 migrant-sending villages between 1972 and 2000. To this aim, they performed both individual-level statistical analyses (namely, event history methods) and village-level analyses (namely, ordinary least squares models) on the empirical data. These models were submitted to a variety of sensitivity and robustness checks concerning variable measurement (in particular, village's social homogeneity) and models' specification (in particular, with respect to unmeasured village's attributes) (ibid., 1919, 1921).

To detect the relevance of macro signaling in generating gender stratification in job-specific skills, Polavieja (2012) combined mathematical and statistical modeling in the following way. First, he formalized investment decisions in jobs requiring learning new skills and formally established that "employees' investment choices will be crucially dependent on their expected duration in the firm" (ibid., 599). He postulated that individuals can estimate their job survival expectations for a given job by looking at others' experiences, namely, "the investment experiences of other women in preceding generations (hypothesis 1), the domestic behavior of men in their proximate social contexts (hypothesis 2), and the fertility behavior of peers (hypothesis 3)" (ibid., 600). Then, he estimated country fixed-effects regressions (with robust standard errors at the region level) on European Social Survey data (2004–2005) concerning 45,600 individuals, nested in 234 different regions and 24 different countries, thus exploiting cross-country and cross-region, within-country variations (ibid., 594, 604, 612), and showed that the three contextual variables were all significantly associated with women's investments decisions in job-specific skills, whereas they had no effect for men (ibid., 616, 619, 621). This between-gender variation was ascertained by estimating the statistical models containing the social interaction effect separately for men and women (ibid., 612). Polavieja (2012) checked that results were robust against various controls for preference heterogeneity, different measures of job-skill specialization, potential regional collinearity, and alternative model's levels specification (ibid., 603, 607, 610, 612, 616–17).

To demonstrate the empirical relevance of attraction/rejection for/of similar/dissimilar others as a social mechanism for homophily in couple formation, Skvoretz (2013: 489, 495, 499–500) combined mathematical and statistical modeling in the following way. He first wrote down equations formalizing the probability of sequential tie formation as depending on (a) the opportunity of encountering similar/dissimilar others as simply the result of the demographic size of the groups considered; and (b) the actors' tendency to prefer similar others and reject dissimilar ones. As the opportunity part of the model only depends on the groups' size, and each of the two individual tendencies are represented by a single parameter, Skvoretz could then exploit empirical cross-tabulations (built from survey data) on marriages,

Gianluca Manzo - 9781789906851

dating, and cohabitation by race, ethnicity, religion, and education, computed to estimate the variation of the values of the two parameters expressing the individuals' bias for similar and dissimilar others across a variety of western countries. Skvoretz (2013: 500, 503) checked the robustness of his results by estimating the model on cross-tabulations artificially generated on the basis of known values of the attraction/repulsion parameters, and also studied model's extension contemplating various forms of inter-group variability in the attraction/repulsion motives.

To study the impact of within- and between-race income inequality on residential segregation by race given the size of race groups, Bruch (2014) articulated agent-based computational models and various multivariate statistical techniques. In particular, she first designed a simplified artificial society with a simplified geography where only artificial agents' residential mobility choices were empirically calibrated, namely through discrete choice conditional logit models estimated on US longitudinal data coming from the Panel Study of Income Dynamics. This first model was used to manipulate experimentally the group size and the level of within- and between-race income inequality (ibid., 1236, 1243). Then, by exploiting Census data, she made the same model more realistic by incorporating within the simulation the geography of three US cities, namely Los Angeles, Atlanta, and Chicago, as well as the temporal evolutions of income inequality and group size in these cities (ibid., 1245–6). Both versions of the computational model pointed to the same possible offsetting dynamics: decreasing between-race income inequality can lead to less isolation for poor blacks, but more isolation (from the black population) of wealthy blacks, the resulting aggregate net effect potentially being unchanged, or even higher, residential segregation (ibid., 1252, 1255). To check if "the model is providing inferences that accurately capture mechanisms at work in real cities" (ibid., 1256), Bruch finally estimated fixed-effects models at the metro area level for 93 major American cities between 1980 and 2000, which modeled segregation level changes in these cities as a function of changes in between- and within-race income inequality, as well as race group size. Traces of the offsetting mechanism that appeared within the simulations were also found in the empirical data described by the statistical models (ibid., 1264–5). Robustness checks concerning variable measurement and models' specification were frequent throughout the analyses (ibid., 1237, 1239, 1248).

To demonstrate that private experiences have different probabilities of traveling through social networks as a function of their stigmatizing potential, Cowen (2014) combined two different sources of empirical data. On the one hand, she designed a tailored internet-based survey investigating the attributes of American adults who were more likely to act as confidantes of women experiencing miscarriage and abortion (ibid., 467, 473, 478). After various controls and weighting procedures to assure national representativeness (ibid., 474), these data were described through logistic regressions modeling the respondents' probability of knowing someone who experienced the two events (ibid., 478). On the other hand, Cowen exploited qualitative data recording women's self-declared motivations for not disclosing miscarriage and abortion: these data came from a qualitative module contained in the Internet-based survey designed by the author herself, as well as from a unique data set containing information on 5000 women who experienced abortion at a US urban abortion clinic (ibid., 475). These qualitative data were used to provide direct empirical evidence that women disclosed their experience preferentially to those from whom they expected support rather than moral disapprobation, a tendency that is stronger as the amount of expected stigma for the event to be disclosed is higher (ibid., 481–2).

Gianluca Manzo - 9781789906851

To document how variation in contextual traces of past (minor) norm violations leads to variation in further (minor) norm violation, thus explaining variations in norm compliance across contexts, Keuschnigg and Wolbring (2015: 103) mainly relied on the experimental approach. In particular, they designed three field experiments (in Munich) where each participant acted in a natural environment and was not aware of being part of an experiment. In the first experiment, littering was used as treatment to signal physical disorder, and subjects were tempted to further littering by finding a nonsense flyer attached to the handlebar of their bike (ibid., 104); in the second experiment, jaywalking was used as treatment to signal social disorder, and the subjects were tempted to same-norm violation by directly observing others' jaywalking at pedestrian lights (ibid., 109); in the third experiment, a lost letter showing a varying amount of money was used as treatment, and subjects were induced to fall in cross-norm violation, i.e. steal the letter after seeing physical disorder (namely, littering) (ibid., 112). All three experiments were set in locations with varying levels of social capital, measured either though tailored survey—for the first experiment (ibid., 106–7)—or through urban district-level data (ibid., 109–11). The statistical significance of differences between experimental conditions was measured through two-side chi-square tests (ibid., 114); a variety of small modifications in the experiment designs were tested to check the results' robustness, and the basic setup of the first and the third experiments was designed to allow replicability of previous experiments (ibid., 104).

To demonstrate that Muslim women may veil for strategic reasons, in particular as a way to signal their piousness when exposed to social contacts perceived as a potential threat to their moral integrity, Aksoy and Gambetta (2016: 792, 795) relied on two national surveys, one from Turkey and one from Belgium (where Muslims are a minority), as well as on a large-scale comparative dataset (the Pew World Muslims Survey) containing information on veiling in 25 Muslim countries in Africa, the Middle East, and Europe. Aksoy and Gambetta (2016: 794, 796) looked into these data with the aim of exploiting variation among women with different levels of religiosity, and among urban areas with various levels of veiling and dispersion in terms of types of veiling. To do so, they employed structural equation modeling (with ordinal probit regressions) treating the respondents' religiosity as a latent variables (ibid., 793, 797), and showed that, while the models' average main effects supported the predictions of the modernization theory (according to which veiling as a manifestation of religious beliefs should be lower among urban, educated, and employed Muslim women), interaction effects making veiling conditional on women's level of religiosity supported the signaling interpretation of veiling: when educated, urban, and middle-class women were highly religious, they veiled more than the average Muslim woman, and they tended to veil more when they lived in a neighborhood where the average level of veiling was lower, and the variety of its dressing expressions was high (ibid., 797, 798, 800). Various sensitivity tests, in particular with respect to missing values imputation and measures of veiling (ibid., 796, 800), were performed to check the results' robustness; secondary qualitative evidence was also described at various points of the analysis to support the mechanism narrative (ibid., 794, 803).

To show that patterns of clustered cultural preferences can emerge from the mechanism of "associative diffusion"—i.e. a sequence of events where actors progressively attach meanings to practices by observing co-occurrences of practices in others' observable behaviors—Goldberg and Stein (2018: 908) overtly followed the principles of agent-based models with "low-dimensional realism" according to which the computational model is used to determine the sufficient condition(s) under which a given macroscopic outcome can

Gianluca Manzo - 9781789906851

appear. In particular, Goldberg and Stein (2018: 899–901) wanted to show that, contrary to what was postulated by many existing works, cultural differentiation can appear even in the absence of selection on practices (so that actors preferentially interact with similar others), and/or network segmentation (which exposes actors only to subsets of practices). To this aim, they designed an artificial society in which (a) each agent possesses a (randomly initialized) vector of preferences for a set of practices, as well as a matrix of associations between these practices (initially all equally associated), (b) agents sequentially interact within random dyadic encounters during which one agent can observe two behaviors of the other agent, and (c) conditionally on interaction, the observer agent updates his matrix of associations between practices and, possibly, their preferences for the observed practices, as long as this contributes to reinforce positive associations between practices (ibid., 908–10). Goldberg and Stein (2018: 913–4) found that this minimal sequence of events is sufficient to gradually lead to stable associations of clustered practices. By further simulating alternative mechanisms (like biased contagion, conformist imitation, and homophily), they also show that none of these mechanism alone (i.e. in absence of network segmentation) is able to generate clusters of cultural practices (ibid., 916–17). As a by-product of these robustness checks, Goldberg and Stein (2018: 920) ultimately established that, if network segmentation is not a necessary condition for the emergence of cultural differentiation when "associative diffusion" is at work, specific network topologies, in particular scale-free-like networks, can generate "crisper" boundaries between clusters of cultural practices.

To document the extent to which adolescents' ethnic identification acts as a driver for ethnic homophily in school-based friendship networks, Leszczensky and Pink (2019: 397) relied on a well-tailored dataset containing three waves of network panel information collected on more than 2000 students in the 5th, 6th, and 7th grades of nine schools in nine towns in the federal state of North Rhine, Westphalia. On this basis they were able to reconstruct "13 grade-level networks consisting of 1,349 students". Leszczensky and Pink (2019: 396–7) studied these networks through "stochastic actor-oriented models for network formation and behavior" (SAOM), which they explicitly compared to agent-based models by noting that SAOM also "assume that observed states are the result of an underlying process that consists of numerous micro-steps in which individuals either form, maintain, or dissolve ties or change their behavior" (ibid., 402). In addition, they claimed (ibid., 403–4) this type of statistical model allows us to assess the relative importance of preferences for similar others by controlling various structural constrains related to the underlying networks and potential social influence (i.e. changes in individual attributes due to tie formation itself). In particular, Leszczensky and Pink (2019: 404–5) are interested in assessing the relative weight of ethnic homophily (i.e. preference for same-ethnicity friends) conditional on *ego* and *alter*'s level of ethnic identification, which they attempted to capture by specifying a two- and three-way interaction term respectively. It turned out that only the three-way interaction effect was statistically significant (ibid., 409), which the author interpreted as evidence that "ethnic homophily is not ubiquitous" but "arises as an interplay of ego's and alter's ethnic identification, requiring strong identification of both parties of the friendship" (ibid., 410). The robustness of this result was checked against various modifications of the variables' operationalization, and the statistical model's specification (ibid., 411).

Finally, to demonstrate that herding behavior can sustain self-correcting dynamics that make status hierarchies with quality/status decoupling only a transient state of the system, van de Rijt (2019) combined mathematical modeling and experiments. In particular, he first formalized

Gianluca Manzo - 9781789906851

discrete choices with uncertainty under social influence to clarify analytically the condition under which low-quality options can stably become more popular, the condition being, according to his model, that the low-quality option is chosen with a probability at least equal to the fraction of the population having chosen that option in the past (ibid., 1472). Then, van de Rijt (2019) considered six experiments that, in various settings and for different types for choices, found that others' behavior was able to distort actors' perception of the quality of the options at hand, thus leading low-quality options to be chosen more frequently than superior ones. By comparing the fraction of choices for the inferior alternative observed in each experiment with the condition for sustainable low-quality dynamics that he identified analytically, van de Rijt (2019: 1473–81) showed that none of these experiments in fact could prove that the fraction of choices for the low-quality options was sufficiently large to generate a stable decoupling of popularity from quality over time. With the aim of creating a sequence of interdependent choices long enough to have the potential to reverse entirely the initial dominance of low-quality alternatives, van de Rijt (2019: 1481) finally designed a novel large-scale, web-based experiment, through Amazon Mechanical Turk, including 5068 subjects, and varying the type of alternatives to be evaluated (ibid., 1483, fig. 2). The analysis of the 27,859 choices recorded showed that "[…] in all seven cases the initial 93% advantage for A was fully eliminated over the course of the experiment and converted into majority support for B" (ibid., 1484). It is worth noting that van de Rijt (2019: 1487) also modified his experimental design to identify conditions under which popularity actually can be stably decoupled by quality, which allowed him to further demonstrate that these conditions are related to structural constrains rather than social influence deriving from herding behavior.

Thus, overall, Merton awarded articles clearly exhibited a large variety of methodological strategies to connect the *explanans* to the *explanandum*. With the exception of Goldberg and Stein (2018), all articles attempted to provide real-world evidence for the proposed mechanism(s), often combining various types of data. While individual-level survey data (in one case also containing ego-centered network information) were the most frequently exploited data, experimental data were at the heart of two studies, as well as qualitative data and historical archival data, respectively exploited by Cowen (2014) and by Kroneberg and Wimmer (2012). As to methods, the tendency also is to combine various methodological approaches. With the exception of Aksoy and Gambetta (2016), all articles combined various forms of formal modeling with different types of statistical models, and/or measurement devices, and/or data collection procedures specifically tailored to the explanatory purpose at hand.

It should be also noted that, when multivariate statistical models constituted the main tool to empirically trace the postulated mechanism(s)—like in Polavieja (2012), Cowen (2014), Aksoy and Gambetta (2016), and Leszczensky and Pink (2019)—the specification of statistical interaction effects is the main leverage to tackle traces of the mechanism in the data. This is an important point because one of the aspects of analytical sociology's original critique of the explanatory power of statistical models was precisely related to the observation that many regression-like models place more emphasis on the evaluation of the relative importance of main additive effects than on designing theoretically motivated statistical interactions between variables. In particular, in his contribution to Hedström and Swedberg's *Social Mechanisms* Sørensen (1998: 244) argued that the specification of these effects is the best way one can exploit statistical models to connect theory and empirics. As "statisticians and statistically oriented sociologists do not like interaction effects in regression models", Sørensen (1998: 245) observed, "[…] we miss out on sociological stories of great interest—how things differ

Gianluca Manzo - 9781789906851

between subgroups and contexts" (see also p. 253). The Merton awarded articles that mainly relied on multivariate statistics clearly followed Sørensen's lesson.

A final remark should be devoted to agent-based computational simulations. As we have seen in sections 1.3 and 1.4, various commentators regretted the increasing emphasis that analytical sociology's *manifestos* put on this method. In particular, more extensively than others, Kalter and Kroneberg (2014: 109) argued that, by "elevating particular modelling or research techniques to foundational principles", and by overemphasizing "particular methodological tools at the expense of others", in particular "empirically calibrated agent-based models", analytical sociology takes the risk of turning into a "mechanism cult". Supposing that the risk is real, Merton awarded articles clearly showed that analytical sociology is far from being exposed to this danger. Agent-based computational modeling is the chief methodology of three out of eleven articles, namely DiMaggio and Garip (2011), Bruch (2014), and Goldberg and Stein (2018); Leszczensky and Pink (2019) can be considered a fourth case in that stochastic actor-oriented models (SAOM) are increasingly considered a specific form of agent-based models, combining empirical calibration and statistical estimation where the *explanandum* is a network configuration (see Snijders and Steglich 2015; Stadtfeld and Amati, this volume). Kroneberg and Wimmer (2012: 191f., 197) admitted the virtues of agent-based computational models to study explanations focusing on the micro-macro connection, but they did not employ it. Thus, while agent-based modeling is clearly part of the toolbox of some of the best articles in analytical sociology, for the moment, this perspective is not dominated by agent-based modeling.

7. MERTON AWARDS' ACTION THEORY AND ANOMALOUS CONCEPTS

We know now that the Merton awarded articles shared a common *modus operandi*. They all started with clearly identified macroscopic patterns or dynamics; they all explicitly pursued the goal of understanding how these system-level properties emerged; they all designed a set of hypotheses relating (collective) actors' motives to the context and the structure of interactions these actors were embedded in; they all argued that the sequential interdependence of those actions and interactions progressively constructed the system-level regularities and/or dynamics of interest; and through various combinations of formal models, experiments, statistical methods, and (with one exception) various types of data, they attempted to document the empirical plausibility of the postulated mechanisms. The Merton awarded articles thus followed in practice the research path progressively designed by analytical sociology's *manifestos* that I discussed in section 1. The main difference concerns the fact that agent-based computational models are not systematically at the heart of the way the model of the postulated mechanism is connected to the *explananda* of interest. In terms of Manzo's (2014: 7–9) synthetic exposition of the analytical sociology's core (see section 1.3), this means that postulates 1 to 4 are always fully realized whereas postulates 5 to 7 are present only in a subset of articles.

In this section, I go back to the way the Merton awarded articles formulated their explanatory mechanism (see section 5, and, for a summary, Table I.2), zoom in on it, and investigate two specific points: first, the extent to which actors' actions were framed within rational choice theories (section 7.1); and, second, the extent to which parts of the proposed mechanism were

Gianluca Manzo - 9781789906851

designed by using "anomalous" elements, i.e. concepts, theories, or broader research perspectives that analytical sociology's *manifestos* either criticized or did not explicitly include within analytical sociology's basic toolbox (section 7.2).

Both aspects seem worthy of special attention. As we have seen in section 1.2, scholars of various confessions criticized analytical sociology's scope for being limited by analytical sociologists' supposedly adherence to rational choice theories of individual action (see, in particular, Abbott 2007a, Gross 2009, Little 2012); symmetrically, scholars who traditionally defended these types of micro-level theories have denied analytical sociology's originality because of its supposed commitment to these same theories of individual action (see, in particular, Opp 2013). Others, although sympathetic to the social mechanism movement, objected to analytical sociology its excessively narrow theoretical framework, in particular one that over-simplifies the "context" in which individual actions take place and downplay the role of "culture" (see Edling and Rydgren 2014, 2016). For these reasons, it seems to me important to assess how articles that were regarded as illustrating analytical sociology at its best locate within this critical space.

7.1 Action Theory

With respect to action theory, analytical sociology's programmatic writings progressively moved toward a more and more explicit pluralistic claim. Hedström and Swedberg's (1998: 11–13) discussed more explicitly methodological individualism than action theory *per se*. They mentioned once the intentional nature of human action (ibid., 12) as well as the fact that social action is typically oriented toward others (ibid., 13) but they did not commit the formulation of action-formation mechanisms to any specific theory of action (ibid., 23). The critique of the variants of rational-choice theory that model actors as computational devices looking for the best information available, and the optimal course of action allowing the maximization of subjective rewards, became explicit in Hedström (2005: 61); the associated critique of "instrumentalism" (i.e. the claim that, although unrealistic, the rationality assumption should not be removed because it may still lead to correct predictions in the aggregate, and/or because it would make formal models easier to study) led Hedström (2005: 61, 65–6) to propose to abandon rational-choice theory as a "general point of departure" for the micro level. Then, the observation that the various forms of "weak" or "subjective" rationality proposed by different authors to increase the realism of rational choice theory raise the problem of opening the way to *ad hoc* and *post hoc* explanations—see, in particular, Manzo's (2012, 2013) analysis of Boudon's theory of reasoned action and Opp's "broad" rational choice theory—progressively moved analytical sociology to question the idea itself of having a general theory of action to be used as a starting point for any explanatory task, no matter whether this theory is only weakly intentionally oriented or more strongly rationally oriented (Hedström and Ylikoski 2014: 65–7; see also Hedström 2021, this volume). Analytical sociology thus arrived to the programmatic statement, summarized in Manzo's (2014: 7) fourth postulate, see section 1.3), that a variety of logics of action—inspired by various versions of rational choice theories, dual-process approaches, heuristic-based action theory, and action theories rooted in experimentally oriented cognitive and social psychology—are in principle all equally legitimate when specifying the micro-level component of a model of given mechanism; the conditions under which it is more realistic to assume this or that logic should be assessed on a case-by-case basis (see Manzo 2014: 21–7).

Gianluca Manzo - 9781789906851

Table I.3's fourth column synthetically shows that the Merton awarded articles fall in three distinct groups with respect to the rationality assumption. Four out of eleven studies—namely, Skvoretz (2013), Bruch (2014), Cowen (2014), and Leszczensky and Pink (2019)—simply did not refer at all to the concepts of rationality, utility, maximization, or strategy to frame the micro-level part of the proposed explanatory mechanisms. Linguistically, this translated to the virtual absence of these terms from the way the *explanans* is described.[8]

Three other articles—namely, Aksoy and Gambetta (2016), Goldberg and Stein (2018), and van de Rijt (2019)—did not contain any explicit statement, comment, or reference to rational action theory, but their *explanans* in practice refer to some forms of the rationality assumption. Aksoy and Gambetta (2016) is the most explicit in this sense in that, although no specific claim on rational choice theory is present, veiling is explicitly interpreted as a double strategic choice, namely "a strategy directed towards the self, and used as a self-binding device against temptation" (ibid., 794), and a strategy directed to others to signal "one's persisting piety" (ibid., 794). Goldberg and Stein (2018: 921–2) and van de Rijt (2019: 1470–1, 1475, 1483) are not that explicit, but both employed the term "utility" or "rationality" in conjunction with the aspect of their *explanans* referring, in the former case, to actors' heuristic of relying on others' co-occurrence of practices to reduce uncertainty about the meaning of those practices, and in the latter case, to actors' heuristic of relying on others' past choices to reduce uncertainty about the quality of a given option.

The last group is made of four articles—namely Keuschnigg and Wolbring (2015), Kroneberg and Wimmer (2012), Polavieja (2012) and DiMaggio and Garip (2011)—which explicitly explained the extent to which they adhere to some forms of rational action theory.

In particular, Keuschnigg and Wolbring (2015: 98) is the only article where the *explanans* is explicitly formulated through the terms and the concepts of rational choice theory: "we offer a rational choice explanation for norm violations in the context of environmental cues for the level of social control," the authors (overtly) claimed. However, as soon as this was posed, Keuschnigg and Wolbring (2015: 98) clarified that the theory that inspired them, i.e. the "Beckerian rational choice model of norm violating behavior" lacks realism in certain situations—"it appears to be unrealistic to assume that actors invest in costly search when breaking minor norms" (ibid., 99). In this case, they argued, "[…] environmental cues signaling social control gain in relevance. Hence, when individual costs and benefits are low it is rational to shorten the lengthy and cumbersome decision-making process by the use of heuristics". We know from section 5 that the heuristic proposed by Keuschnigg and Wolbring is that the traces of past behaviours in the actors' proximate environment are assumed to be the clue to form subjective beliefs on the probability of conviction. This is a form of rational imitation of others' behaviors. Social embeddedness is the second enrichment of the standard rational choice framework proposed by Keuschnigg and Wolbring (2015: 98, 102–3). They argue that the perceived level of social capital in a given neighborhood is also taken into account by a rational actor when deciding whether or not to violate a norm. In particular, if the social

[8] One may want to consider Bruch (2014: 1237) as a borderline case in that she employed the term "utility" to express the intensity of the preference that a given actor associates to each residential neighborhood where she could potentially move. I am reticent to do so because the word "utility" is employed by Bruch only when she is presenting the dependent variable of discrete choice models to be statistically estimated, but she did not formulate any explicit behavioral interpretation of these models in terms of individual utility maximization. Thus I suspect that Bruch wanted to remain agnostic about the substantive meaning in action terms of the proposed logistic regression models.

Gianluca Manzo - 9781789906851

capital is perceived as low, the signaling force of traces of past deviating behaviors should be weaker—norm violations are perceived as common in this context anyway; vice versa, when the social capital is perceived as high, the signaling force of traces of past deviating behaviors should be high too—these traces suggest that norms can in fact be violated.

Kroneberg and Wimmer's (2012) study is also very close to rational choice theory. As we have seen, the *explanans* combines an exchange model of preference formation, where actors are assumed to prefer political alliances that maximize their control over political and economic resources (ibid., 190), with a non-cooperative sequential game of alliance formation. However, Kroneberg and Wimmer repeatedly emphasized the differences that exist between their approach and basic rational choice theory. First, they clarified that, in fact, they only assumed "[…]—according to a weak version of standard rationality assumptions—that actors did X because they wanted to achieve Y, which is often the consequence of doing X […]" (ibid., 182). Second, with respect to the preference formation model, they emphasized that they went beyond a "purely instrumental approach" (ibid., 180), "purely utilitarian logic" (ibid., 187), and "instrumental interests" (ibid., 190), in particular, by adding "[…] an additional component into the model by letting actors consider with whom they can identify on the basis of cultural similarity when choosing alliance partners" (ibid., 180, see also pp. 187, 190 and 196). Third, with respect to the negotiation model, Kroneberg and Wimmer (ibid., 197) insisted on the fact that they "go beyond standard game-theoretic models that do not consider the unequal distribution of resources and treat the preferences of its players as exogenous—a major point of critique by scholars both sympathetic (Elster 2000) and unsympathetic to rational choice theory (Somers 1998)".

As to Polavieja (2012), his proposed *explanans* for women's under-investment in jobs requiring job-specific skills started with a representation of an actor's decision on the job market resulting from cost-benefit calculations. However, Polavieja (2012: 596) made it explicit that his goal is to show that these calculations "[…] are always and crucially influenced by the social structure in which individuals are embedded […]". In particular, given the uncertainty surrounding the evaluation of job survival probabilities, the actions and the outcomes experiences of other actors in the actor's proximate social context provide clues to assess these probabilities: rational expectations can be formed through rational imitation (ibid., 603). Thus, similarly to van de Rijt (2019), Goldberg and Stein (2018), and Keuschnigg and Wolbring (2015), rational action finally takes the form of a simple heuristic: "if you are uncertain, follow others' behavior". Polavieja (2012: 622) explicitly acknowledged that "The idea that uncertainty and information constraints make individual choices more context dependent can be met by rationality assumptions of all strengths" (for a discussion, see Boudon (1996) and Goldthorpe (2000, Ch. 6)).

Finally, if DiMaggio and Garip's *explanans* (2011: 1888) is explicitly based on "goal-directed individual actions", they also made equally clear that they "[…] do not presume that choices are 'rational', either phenomenologically or in effect. People may adopt courses of action unreflectively, or they may adopt them on the basis of poor information about costs, benefits, and risks" (ibid., 1893, ft. 5). Again, action is represented as driven by a simple heuristic, namely to follow prior actions accomplished by the focal actor's direct contacts (ibid., 1888).

Thus, overall, the Merton awarded articles are clearly all committed to an actor-centered perspective. However, they do not uniformly rely on assumptions of rational choice theories to design the actor-related part of their explanatory machinery. Actors are never depicted as having optimal information and looking for subjective utility maximization. When individuals'

Gianluca Manzo - 9781789906851

actions are explicitly or implicitly depicted as rational, rationality is conceived as variously limited by the context, and usually take the form of heuristic-driven behaviors where the behaviors of others are exploited to reduce uncertainty about the quality, the risk, the attractiveness, or the meaning of the options under consideration.

7.2 "Anomalous" Elements

Table I.3's fifth column synthetically shows that the explanatory machinery of the Merton awarded articles also incorporated "anomalous" elements, i.e. concepts, theories, or broader research perspectives that analytical sociology's *manifestos* either criticized or did not explicitly include within analytical sociology's basic toolbox.

In this respect, the notions of "structure" and "context" are a first interesting case to be considered. Edling and Rydgren (2014: 1) admonished analytical sociology's *manifestos* for reducing social structures to "structures constituted of interacting individuals" as well as for giving priority micro-to-micro over macro-to-micro mechanisms. They urged analytical sociologists to consider that, in fact, "individuals' positions in the social structure, culture, as well as their identification with groups, social categories, and collectives, shape their beliefs and desires, and actions, in important ways" (Edling and Rydgren 2016: 1139).

Hedström's (2005: 17–18) critique of Blau's analysis of organizational differentiation—which Hedström saw as "a direct precursor of his later so-called macro-sociological theory of social structure"—can partly explain this type of objections. In particular, Hedström remarked that Blau's style of theory construction equated the explanation of a fact to an operation of deduction of this fact from a hierarchy of propositions characterized by a decreasing level of abstraction and, given that these propositions are claimed to relate structural elements without making reference to individual-level properties, this type of theory construction is at odds with analytical sociology's ambition to open correlational black boxes (ibid., 19, 32).

Interestingly, among the Merton awarded articles, two studies—namely, Skvoretz (2013) and Bruch (2014)—framed their own contribution precisely within the context of Blau's theory of social structure or, more precisely, his analysis of how the composition of a given population impacts its inequality patterns. In particular, Skvoretz's (2013) model of homophily generation conceived individuals' attraction for similar others (and rejection of dissimilar others) as tendencies that bias the chance of encountering similar (dissimilar) others due to the simple sizes of the groups at stake. Skvoretz (2013: 487) overtly stated that this second aspect, i.e. the intuition that "some ties between similar others will form by chance and that the rate at which this occurs depends on the population distribution along the attribute of interest", is "[…] central to Blau's 1977 theory of social structure". However, Blau's central concern was "[…] how changes in overall population composition could create more or less opportunity for intergroup relations", noted Skvoretz (2013: 507); he wanted instead to understand more specifically the possible sources of inbreeding homophily (given baseline homophily). Skvoretz (2013: 487) noted again that Blau actually considered inbreeding homophily a "serious topic—because intergroup relations are necessary for societal level integration, yet inbreeding homophily inhibits their occurrence"—but he did not explore its underlying mechanism. Similarly, Bruch (2014: 1221) explicitly related her investigation of the interplay between race group's sizes, between- and within-race income inequality, residential choices, and residential segregation to one of the aspects of Blau's theory of social structure, namely the correlation that exists between levels of "consolidation" (i.e. the strength of the correlation

Gianluca Manzo - 9781789906851

between different attributes of the social structure) and social differentiation. Bruch (2014: 1222, see also p. 1228) noted that "[…] while Blau's original framework is compelling, it conflates two dimensions of the correlation between social attributes: within-attribute variation versus between-attribute variation", and she declared that her own goal is to "[…] extend Blau's work to investigate how choices about social grouping based on one attribute can exacerbate or attenuate segregation on another correlated attribute. The specific application is the role of racial and economic factors in generating patterns of racial residential segregation".

Thus, although both Skvoretz (2013) and Bruch (2014) seemed to ignore the skepticism that analytical sociology's *manifestos* expressed on the possible holist bias of Blau's theory of social structures, both studies in fact made some of the aspects of this theory compatible with well-specified social-mechanism-based explanations by turning structural elements into one of the parameters of individual choices, namely actors' opportunities.

"Social identity" is a second concept that some have argued being of high explanatory relevance while not being explicitly covered by analytical sociology's *manifestos*. In particular, with the aim of saving the predictive power of theories of rational action that equates "rationality" to "reason-based" behavior, Manzo (2012) argued that analytical sociology should pay more attention to individuals' social identities because identity often acts as a "reason trigger", i.e. an individual-level factor that tends recurrently to activate sets of specific reasons in specific contextual settings. Edling and Rydgren (2014) endorsed a more radical view. They regarded analytical sociology as still excessively biased toward intentional explanation (ibid., 2, 5) and argued that one way to counteract this tendency would be to admit that "social actors also act out of identity, which comprises shared action potentials neither reducible to individual interests nor, in fact, to intentionality" (ibid., 10). They urged analytical sociology to pay special attention to "overlapping system of identifications" in which the individual's identity is instantiated and to study the conditions under which this or that social category acquires cognitive salience (ibid., 11–14).

Interestingly, among the Merton awarded articles, two studies—namely, Kroneberg and Wimmer (2012) and Leszczensky and Pink (2019)—overtly put actors' social identity at the heart of their explanatory machinery. In particular, Kroneberg and Wimmer (2012: 184) related their mechanism of alliance formation to identity formation and, by invoking "a long line of research in social psychology", they explicitly stated that "we assume that actors who enter into an alliance with each other will also develop a shared identity over time". This process is partly explained by the capacity of some actors to control cultural institution (like school or print media), thus allowing them to disseminate more effectively their "[…] vision of the legitimate divisions" of society, to paraphrase Bourdieu" (ibid., 194). These are the elements that brought Kroneberg and Wimmer to note that their model "[…] avoids the solipsistic bias that characterizes many formal models of individual identity choice", and that "[…] the model contributes to a growing literature that attempts to translate certain strands of the constructivist literature on nationalism and ethnicity into a formal modeling architecture […]." At the same time, Kroneberg and Wimmer (2012: 193, 194) took the precaution of explaining that their model cannot be equated to "elite manipulation arguments" because all actors are represented to have control over some resources (masses, for instance, control "military support") so that "group formation and collective identities cannot be directly generated by force". Similarly, Leszczensky and Pink's (2019: 395) referred to social identity theory (in the tradition of Tajfel's work, one of the references also invoked by Kroneberg and Wimmer (2012: 398)) to better specify how individual preferences for similar others operate in the formation of

Gianluca Manzo - 9781789906851

ethnic homophily. In particular, their hypothesis that "ethnic identification" should be seen as a modifier of preferences for same-ethnicity friends is directly deduced from "social identity theory" that "[…] links the strength of ethnic identification to individual preferences" (ibid., 398) and "states that people strive for positive social identities, which they obtain by favorable comparisons of their own group with other groups" (ibid., 399). More particularly, Leszczensky and Pink's (2019: 398) built on "[…] the affective dimension of social identity, understood as the "emotional significance" individuals attach to their group membership […]", which they considered the most important ingredient of ethnic identity when one wants to understand preferences for the in-group. Leszczensky and Pink (2019: 398) emphasized that they did not only rely on social identity theory, but extended this perspective through their network perspective focusing on "the interplay of two (potential) friends' ethnic identifications" (ibid., 411).

Finally, let me consider how different views of social interactions themselves penetrated the Merton awarded articles. In this respect, Erikson's (2013) typology of the theoretical frameworks underlying social network analysis turns to be very useful. In particular, she demonstrated that this field is animated by two broad perspectives that she synthetically labelled "formalism" and "relationalism". While the former sees "networks as pipes", i.e. structures of entities with stable and given identities constraining those entities' opportunities, beliefs and ultimately actions, the latter sees networks as culturally constructed and constantly moving objects with unstable boundaries and identities. Erikson (2013: 222–3, 236) noted the proximity between "relationalism", pragmatism, and cultural sociology and, more recently, she rightly added that analytical sociology has tended to treat social network analysis more from a "formalist" viewpoint than from a "relationalist" perspective (see Erikson 2017: 292–3).

Interestingly, among the Merton awarded articles, two of them—namely Cowen (2014) and Goldberg and Stein (2018)—clearly endorsed a "relationalist" perspective and overtly put subjective perceptions of social networks at the heart of their *explanans*. More particularly, when Cowen (2014) proposed to explain the temporal stability of attitudes toward abortion in terms of women's choices of selectively disclosing their private experiences as a function of the expected reaction of the potential confidante (using political attitudes as a clue to formulate such expectation), she explicitly claimed that her mechanism reverses the dominant perspective on the relation between actors' attitudes and networks: "Scholarship has already shown that social networks shape attitudes; I show that through secret keeping and sharing, attitudes shape individual experiences of social networks as well" (ibid., 484). In particular, she wanted to show that selective disclosure sustain a "self-fulfilling illusion"—i.e. "[…] those who are opposed to a given secret are less likely to hear of it even if it exists in their social vicinity" (ibid., 485)—so that homophily on opinions and behaviors does not exist in *ego*'s local network but *ego* perceives it—"[…] ego can exist in a network where people do not hold a person's attitudes or behave as the person would like but the person mistakenly perceives that they do". It is precisely this falsely subjectively perceived lack of diversity that create the basis for the absence of opportunities for belief change and, consequently, opinion stability at the macro level.

Similarly, Goldberg and Stein's (2018) computational model of how clusters of cultural practices emerge in the absence of pre-existing network segmentation is overtly framed within a meaning-centered, cognitively rooted view of social structures:

> In the theory we propose, rather than beliefs or behaviors per se, the cultural elements being transmitted between individuals are perceptions about which beliefs or behaviors are compatible with

Gianluca Manzo - 9781789906851

> one another […]. Ultimately, our model turns the causal arrow in conventional accounts of social contagion on its head: we show that differentiation can emerge through the complex ways culture is cognitively represented and acted upon by individuals. (ibid., 898–9)

Goldberg and Stein (2018: 907) explicitly claimed that, whilst "[…] students of cultural diffusion have overwhelmingly left meaning out of their epidemiologically inspired models […]", their alternative model of associative diffusion describes how "[…] arbitrary cultural meanings emerge and become consensually accepted through social interaction" and, in this sense, it describes a process of "social construction (Berger and Luckmann 1967)": "People do not simply imitate others' behaviors; rather, they learn, by observing others, what these behaviors mean". For this reason, Goldberg and Stein's (2018: 922) regarded their study as a contribution to both sociology of network and diffusion—"[…] our associative model shifts focus from the diffusion of practices to the diffusion of interpretation—and to sociology of culture—[…] we demonstrate how meaning arises through the process of associative diffusion".[9]

Thus, in six out of eleven Merton awarded articles, theories, concepts, or broader perspectives that were either criticized, or not explicitly considered as a part of the initial basic toolbox of analytical sociology, are explicitly employed to build well-specified mechanism-based explanations. The common point of these articles is that these "anomalous" elements are always conceptually unpacked so that the concept at hand—be that of structure, identity, or meaning—is never used as a pure label. This is achieved either through formalization, thus making the meaning and content of the concept clear and replicable by others (this is the case of Kroneberg and Wimmer, Skvoretz, Bruch, and Goldberg and Stein), or empirically measured through well-tailored data collection procedures (this is the case of both Cowen and Leszczensky and Pink's use of questionnaires to quantify stigmatization and ethnic identification respectively).

8. SUMMARY OF THE ANALYSIS

The main goal of this chapter was to provide a first comprehensive assessment of what analytical sociology looked like when scholars, inspired by its principles and methods, focused on solving specific substantive problems. To do so, I considered the eleven journal articles that have been awarded the Merton paper award (or an honorable mention for it), the prize the *International Network of Analytical Sociology* created in 2012 with the explicit intent of recognizing works that best illustrate analytical sociology's research practices. For this reason, although specific, this sample seems a natural starting point for a first assessment of the extent to which analytical sociology has so far practiced what it has preached. The members of the

[9] Clear traces of a constructionist view of social structure can also be found in Leszczensky and Pink's (2019) study of mechanisms driving ethnic homophily. As we have seen, these authors deduced their main hypotheses from social identity theory. This same theory led them to regard ethnic groups from a subjective perspective—"[…] most studies on ethnic homophily treat ethnicity (or race) as a seemingly objective category rather than a potential source of social identity whose *subjective* importance varies both between and within individuals over time" (ibid. 395). They also explicitly argued for treating perceived group belonging as a relational process—"[…] our relational social identity approach calls scholars to explicitly consider what being a member of a social group actually means […]; doing so requires going beyond the individual level: a deeper understanding of how ingroup preferences unfold and structure diverse networks can only be obtained by considering the strength of ingroup identification of all parties involved in a social relationship" (ibid., 415).

Gianluca Manzo - 9781789906851

various prize committees (see Table I.1) were certainly sympathetic to the analytical sociology movement—if only because they were selected among participants to the INAS annual meetings or were suggested by those participants—but they clearly have different substantive, theoretical, and methodological personal *agenda* so that it was not *a priori* obvious whether a stable consensus on what analytical sociology is would appear across the prize's years.

To evaluate in a principled way the extent to which this was the case, the first section of the chapter was devoted synthetically to reconstruct the evolution of analytical sociology's principles and methods through analytical sociology's *manifestos* over the last twenty-five years. This preliminary analysis showed that analytical sociology progressively evolved toward a clearly codified research strategy organized around a well-defined sequence of research heuristics and steps. The set of seven postulates proposed by Manzo (2014: 7–9) to summarize this evolution of analytical sociology's programmatic *agenda* was then adopted as a benchmark in order to assess which principle and method was present, and how regularly, within the eleven Merton awarded articles.

This analysis generated three main general results concerning the general *modus operandi* of the Merton awarded articles and two additional observations concerning some specific aspect of this *modus operandi*.

General result 1—All Merton awarded articles started with a clear definition (typically provided in the article's introductory section) of a well-defined macro phenomena that they had the ambition to explain. These macro phenomena concerned an impressive range of structural properties and dynamics and could equally be a point of equilibrium of a social system, transitions from one equilibrium to another, dynamic flows, associations between contextual features and distribution of individual's attributes or choices, associations between population distributions, or properties of configuration of social ties. In each case, the article's goal was to explain the system-level patterns and/or dynamic, and explaining meant reaching an understanding of how the *explanandum* of interest was produced.

General result 2—To satisfy this explanatory goal, all Merton awarded articles designed a set of hypotheses, i.e. the *explanans*, systematically and consistently combined the three following ingredients: (a) a description of the specific pathways through which the context in which actors act shapes their beliefs, preferences, and/or opportunities; (b) the principles underlying actors' actions; and, (c) how action interdependency progressively generates the macro outcome of interest. In this sense, each awarded article designed a mechanism, i.e. a detailed description of a multi-level dynamic chains of events that, through specific action's logics and interdependence structures, is likely to lead to the system-level states and/or changes of interest.

General result 3—All Merton awarded articles attempted to prove that the *explanans* could plausibly and realistically lead to the *explanandum*. With only one exception, this was done by studying (most often) individual-level survey data, but also experimental, qualitative data, and/or historical archival data. As to methods, most of the articles combined various forms of formal modeling (including agent-based computational simulations) with different types of statistical techniques and/or measurement devices and/or data collection procedures specifically tailored to the explanatory purpose at hand. When multivariate statistical models constituted the main tool to tackle traces of the mechanism in the data, the priority was given to the specification of statistical interactions over the analysis of additive marginal effects.

Considered together, these three general results showed that all the eleven Merton awarded articles, in practice, followed analytical sociology's theoretical principles concern-

Gianluca Manzo - 9781789906851

ing how to define the research question, what constitutes a relevant *explanandum*, how to explain it, and what the structure of the explanation should basically contain. These elements correspond to postulates 1 to 4 of the analytical sociology's core research strategy, as synthetically defined by Manzo (2014: 7–9). In contrast, methodological postulates 5 to 7, emphasizing the power of agent-based computational modeling as a pivotal tool to reconstruct the *explanandum* in a generative way, only found partial application. While four out of the eleven Merton awarded articles exploited this approach to create the *explanandum*/*explanans* connection, the majority of them relied on more traditional statistical and, in two cases, experimental methods.

As anticipated, the careful investigation of the Merton awarded articles also generated two more specific observations on how analytical sociology was realized in substantive works.

Specific observation 1—Although all the Merton awarded articles are clearly committed to an actor-centered perspective, they did not uniformly rely on assumptions of rational choice theories to design the actor-related part of their explanatory machinery. Actors were never depicted as having optimal information, or looking for subjective utility maximization. Rationality was always conceived as variously limited by the context and usually took the form of heuristic-driven behaviors, where the behaviors of others were presented as a solution to reduce the uncertainty surrounding the evaluation of the options under consideration.

In this sense, the Merton awarded articles instantiated analytical sociology's call for action theory pluralism—i.e. the claim that, if we agree on the goal of achieving realistic explanation, then there is no theoretical nor empirical reason to consider a given theory of action as the obliged starting point for the model building stage (see Manzo 2014: 21–7; Hedström and Ylikoski 2014; see also Hedström 2021, this volume).

Specific observation 2—A consistent proportion of the Merton awarded articles (namely, six out of eleven) was capable of integrating within the proposed well-specified explanatory machinery, theories, concepts, or broader perspectives that analytical sociology's *manifestos* either criticized or did not explicitly consider part of the initial basic toolbox of analytical sociology—like Blau's theory of social structure, social-psychology works on social identity, and meaning-centered perspectives on network analysis, inspired by pragmatist and cultural sociology (see also Barbera 2021, this volume).

In this sense, the Merton awarded articles demonstrated in practice that analytical sociology can face the challenged formulated some years ago by Edling and Rydren (1994: 10), who argued that

> in order to prove itself, analytical sociology will have to take seriously some of the "soft" ideas that are at the (mainstream) core of sociology such as culture and identity. We claim that there is no lack of sources of inspiration in the literature, and that culture and identity can be dealt with within the confines of analytical sociology and in fact serve to improve it.

8.1 Five Guidelines and the Present *Research Handbook*

To conclude, I would like to exploit the results and observations I just summarized to formulate five guidelines for the further development of the analytical sociology's research program. The Merton awarded articles, in fact, also contained elements that can be used to pinpoint aspects of the practice of analytical sociology on which we could/should continue to work. In particular, I detected five important lines of evolution for the future, to which the present *Research Handbook* intends to contribute.

Gianluca Manzo - 9781789906851

Guideline 1—Be explicit about whether your substantive investigation is related to the analytical sociology program and, if so, be clear about what theoretical and methodological principle(s) are you following, challenging, or refining.

While the Merton awarded articles followed the theoretical principles of analytical sociology's *manifestos* and, to a large extent, the methodological ones (the partial exception being a limited use of agent-based computer simulations), the majority did not explicitly refer to analytical sociology's *manifestos* to frame their substantive investigations and methodological recipes. All the articles widely relied on concepts, models, and typical mechanisms that analytical sociology's *manifestos* recurrently regarded as founding pieces of the analytical sociology's toolbox, but explicit and systematic use of the label "analytical sociology" was rare and systematic confrontation with the programmatic texts was virtually absent. While this choice can be regarded as a virtuous sign that, contrary to what critics sometimes claimed (see Chattoe-Brown 2014), analytical sociology is not a "self-regarding club," one should appreciate that this lack of explicit programmatic elaboration on the extent to which a given piece of work adheres to a given research program can have two potentially damages. First, substantively, this slows down analytical sociology's capacity to refine and enrich its own theoretical and methodological principles. Second, institutionally, not referring to analytical sociology's research program when one is using it diminishes analytical sociology's visibility and distinctiveness, thus further reducing the probability that others may want explicitly to refer to this perspective when, in fact, using it. Making explicit the adherence to a research program emphasizing a loop between clarity, precision, explanations fighting black boxes, formal models, and innovative data collection procedures seems especially important in a time where the legitimacy of the project of sociology as a science is strongly challenged by competing intellectual projects variously equating sociology to academic-based political activism (see Turner 2019; Romero 2020).

Guideline 2—Be explicit about how you use the concept of mechanism and how your own mechanism(s) relate to those that have been previously elaborated.

Each Merton awarded article provided a detailed description of a multi-level dynamic chains of events sustained by specific contextual constrains, action's logics, and interdependence structures. In this sense, they all escape without ambiguity to what Kroneberg and Kalter (2014: 100–2) called "mechanism talk"—i.e. various types of poorly explicated explanatory machineries like "pseudo mechanisms" (where "mechanism is just another word for the phenomenon of interest"), "ad hoc mechanisms" (descriptions, story-telling or interpretation lacking "evidence that the processes indeed show some *generality* and do indeed *apply* under the given conditions"), "rudimentary mechanisms" (reference to generic concepts without "further details on the bridge hypotheses, the action theory, and the transformation rules" or reference to "intervening variables that are mistakenly seen to 'explain' the presumed causal effect of an independent variable on a dependent one"), or "partial mechanisms" (simple references "to broad theoretical models or approaches"). At the same time, I observed that the majority of articles employed the term "mechanism" without defining it explicitly, and did not refer to typologies of mechanisms proposed by analytical sociology's *manifestos*. One should appreciate that this constitutes a considerable threat to knowledge accumulation by making the comparison of mechanisms across studies more difficult. A concrete example of this problem is given by the six (out of eleven) Merton awarded articles that postulated various forms of

Gianluca Manzo - 9781789906851

dependence between the focal actor's choice and prior choices of actors the focal actor can see, either through their social ties or through the aggregate. Only one article explicitly referred to the concept of "rational imitation"; none of them explicitly located this specific part of the larger explanatory mechanism postulated within analytical sociology's typologies of how social interactions may impact the various actions' components. This does not facilitate mechanism comparison, but obliges the reader to guess the specific channel (opportunity, preferences, or desires?) and the specific mode (rational or automatic?) through which the dependence across actors' actions is assumed to operate; this reduces the possibility to see that similar mechanisms are shown to be at work in different contexts and for different macroscopic phenomena.

Guideline 3—Be explicit about the type of action theory that informs the micro-level part of the proposed mechanism and, more particularly, be explicit about whether some form of rationality is postulated for individuals' actions.

The Merton awarded articles uniformly adopted weak forms of rationality and/or various forms of heuristic-based behaviors. This is in line with analytical sociology's skepticism about the possibility of using the same action theory as the general starting point for any modeling exercise. At the same time, only a minority of articles explicitly explicated the extent to which the micro level part of the proposed mechanism fits into rational action theories. Again, this lack of explicit theoretical elaboration threatens knowledge accumulation. In particular, it slows down the empirical understanding of the conditions under which strong and weak rationality assumptions hold and, more generally, when different types of action logics are more likely to be realistic. I observed that several Merton awarded articles explicitly identified the level of uncertainty as a crucial aspect to determine how strongly rationality can be assumed to operate. Others are silent on this point, however. For instance, when Cowen (2014) predicated selective disclosure on women's desire to avoid social stigma, was she assuming that women act strategically to preserve their reputation? Or, are women supposed to react emotionally and out of reflection? Or, still, when both logics are present in this case, what type of combination of "hot" and "cold" reasons is the modeler postulating? This is the type of question that we could progressively answer empirically if the assumptions one makes at the micro level were explicitly discussed in connection with the empirical phenomenon under investigation. Being clear on what program of action theory, "weak" or "strong" (see Kroneberg and Tutic, this volume; see also Hedström, this volume), one is following, and making the effort of unpacking broad concepts like uncertainty (see Ermakoff 2017; Ermakoff 2021, this volume) should help to go further into this direction.

Guideline 4—Be explicit about the concepts and theoretical perspectives that you bring into the analytical sociology's research program, about why you want to do so, and about how you do it.

Several Merton awarded articles built part of their explanatory machinery by relying on theories, concepts, or broader theoretical perspectives that were either criticized or not explicitly considered as a part of the initial basic toolbox of analytical sociology by analytical sociology's *manifestos*. This fact can see as the strongest indicator that, contrary to what some critics claimed (see Chattoe-Brown 2014), analytical sociology is far from being a "self-regarding club". When specific substantive problems are at stake, scholars following analytical sociology's principles turned out to be pragmatic: they picked up theoretical ingredients that seemed

Gianluca Manzo - 9781789906851

empirically adequate to deepen the explanation. Merton awarded articles that engaged with the concepts of structure, of social identity, or of meaning did so under a specific constrain: the concept is never used as a broad covering label. Either through formalization or well-tailored data collection procedures, the content of the concept, and its implications are clearly specified (see also Hedström, this volume). Once again, however, this operation of theoretical enrichment is more implicit than explicit. None of the Merton articles that built on concepts or theoretical perspectives that analytical sociology's *manifestos* did not cover initially did so with explicit reference to analytical sociology. This has two potentially undesirable consequences. On the one hand, it makes less visible analytical sociology's theoretical openness; on the other hand, it makes less visible analytical sociology's capacity to make transparent theoretical constructs that social theory has traditionally tended to treat as too rich to be formalized and/or quantitatively measurable. Analytical sociologists who intend to bring "anomalous" concepts or theories within analytical sociology should do this explicitly and systematically. An effective example to follow is Jaeger and Breen (2016) where Bourdieu's theory of cultural reproduction is formalized with the overtly displayed goal of making it clearer and equipped with solid micro foundations, and then challenged through empirical data.

Guideline 5—Be explicit about the extent to which each component of the proposed explanatory mechanism can be really studied through the data and the methods you decide to employ in order to connect the explanans *to the* explanandum.

All Merton awarded articles attempted to prove that the *explanans* could plausibly and realistically lead to the *explanandum* through various combinations of formal modeling, statistical techniques, large-scale datasets, and/or original data collection procedures specifically tailored to the explanatory purpose at hand. However, the specific methodological procedure adopted by some did not allow to document empirically all the ingredients of the proposed explanatory mechanism. While all the Merton awarded articles conceptually formulated a mechanism clarifying how the context constrains actors' action, the principles driving actors' actions, and how the actors' actions are interdependent, some articles relied on methods that could not explicitly represent this last component of the proposed mechanisms. Analytical sociology's *manifestos* progressively emphasized agent-based computational models as a flexible tool to formalize how interdependently constrained behaviors construct a given macroscopic outcome sequentially and dynamically. Some Merton awarded articles demonstrated that in field experiments, large-scale experiments run through the web, and individual-level longitudinal data combined with ego-network information can also be exploited to create, and document empirically, an aggregative dynamic, thus clearly indicating under which conditions dynamic interdependent actions can generate a given macroscopic fact or trend. In contrast, the Merton awarded articles that mainly relied on cross-sectional, individual-level survey data analyzed through multivariate statistics—think in particular of Polavieja (2012), Cowen (2014) and Aksoy and Gambetta (2016)—were strong in documenting empirically the macro-to-micro and the micro-to-micro part of the proposed mechanisms, in particular showing how various contextual and individual-level attributes changed the probability of certain individual-level outcomes, but they were incapable to formalize, nor document, whether the combination of those outcomes in fact led to the macroscopic outcome of interest, be that gender inequality, patterns of opinions, or clusters of religious behaviors. Makovi and Winship's (Figures 20.1 and 20.2, this volume) representation of Aksoy and Gambetta's and Cowen's explanatory mechanisms

Gianluca Manzo - 9781789906851

through the variable-centered perspective of directed acyclic graphs (DAGs) powerfully illustrate this point in very effective graphical terms. The danger of this is proposing verbal narratives involving the micro-macro links without explicitly documenting under which conditions the postulated aggregation process can lead to the system-level outcome of interest. Being more explicit about what component of the proposed explanatory mechanism of the method we can represent and document would help analytical sociology identify potential gaps in the proposed explanations, thus progressively increasing explanatory depth itself.

The present *Research Handbook* has the ambition to contribute to the further development of analytical sociology along these guidelines. In particular, the chapters included in the first part of the book aim to further clarify analytical sociology's principles, paradigmatic antecedents, and borders. In this sense, these chapters share the common goal of contributing to fulfill *guidelines 1 and 4*. This is done in three different, but complementary, ways. First, some of analytical sociology's fundamental conceptual tools, like the Coleman boat (Chapter 1) as well as some aspects of analytical sociology's historical roots like analytical Marxism and analytic philosophy (Chapter 2) and structural individualism (Chapter 3) are reconsidered. Second, similarities and differences with potentially closed research fields, like complexity science (Chapter 4) and social stratification research (Chapter 5), are explored. Finally, a more explicit comparison is performed between analytical sociology and theoretical perspectives like critical realism (Chapter 6), cultural sociology (Chapter 7), and pragmatism (Chapter 8) that are in principle more distant from analytical sociology.

The chapters included in the second part of the book aim more specifically at increasing knowledge accumulation within analytical sociology, and to favor mechanism comparison. In this sense, they all pursue *guideline 2*. They do so in two different but complementary ways. On the one hand, they overview how analytical sociology has treated basic conceptual building blocks for designing models of mechanisms, in particular individuals' actions (Chapter 9), dyadic interactions (Chapter 10), network structures (Chapter 11), norms (Chapter 12), larger contexts (Chapter 13) as well as generic elements shared by all these building blocks like contingency and randomness (Chapter 14). On the other hand, some chapters systematize what we know about generic mechanisms that have been especially studied by analytical sociologists, which include cumulative advantage mechanisms (Chapter 15), phenomena of network amplification (Chapter 16), and complex contagions (Chapter 17). By providing conceptual coordinates for action theories, Chapter 9 also contributes more specifically to *guideline 3*.

Finally, the chapters included in the third part of the book have the goal of increasing analytical sociologists' awareness of how various types of data and methods can specifically contribute to formalize and document empirically social mechanisms. In this sense, they all have the common goal of contributing to *guideline 5*. In particular, as to data, I gave priority to information that has been less frequently exploited by analytical sociologists, in particular historical and qualitative, as well as numerical archival data (Chapter 18), and various types of digital traces (Chapter 19). As to methods, I wanted to cover both methods putting a strong emphasis on causal inference like mediation analysis (Chapter 20) and experiments (Chapter 21) and methods focusing on dynamics and generativity like game theory (Chapter 22), network models (Chapter 23), and agent-based simulations (Chapter 24). To stress the complementarity of these resources to represent and document a given social mechanism, a claim for a many model approach is finally examined in Chapter 25.

Gianluca Manzo - 9781789906851

In conclusion, Hedström's *Coda* provides a personal reflection on various aspects of the past and the future of analytical sociology by focusing especially on whether, and, if so, under which conditions intentional explanations should still be regarded as a crucial step for the construction of empirically plausible models of social mechanisms, thus provocatively contributing to *guideline 3*.

REFERENCES

*Articles which received the Merton Award for the best paper in analytical sociology

Abbott A. (2007a), "Mechanisms and Relations", *Sociologica*, 2.

Abbott A. (2007b), "Mechanisms and Relations: A Response to the Comments", *Sociologica*, 2.

*Aksoy O., and D. Gambetta (2016), "Behind the Veil: The Strategic Use of Religious Garb", *European Sociological Review*, **32,** 6: 792–806.

Barbera F. (2012), "Meso-Level Mechanisms and Micro-Level Foundation", *Sociologica, Italian Journal of Sociology On Line*, 1. https://doi.org/10.2383/36899.

Boudon R. (2012), "'Analytical Sociology' and the Explanation of Beliefs". *Revue européenne des sciences sociales*, **50**, 2: 7–34.

Bouvier (2018). "Intentional, Unintentional and Sub-intentional Aspects of Social Mechanisms and Rationality: The Example of Commitments in Political Life", in G. Bronner and F. Di Iorio (eds), *The Mystery of Rationality. Mind, Beliefs, and the Social Sciences*, Cham (Switzerland): Springer, Ch. 3.

*Bruch E. (2014), "How Population Structure Shapes Neighborhood Segregation", *American Journal of Sociology*, **119**, 5: 1221–78.

Chattoe-Brown E. (2014), "Review of Analytical Sociology: Actions and Networks", *Journal of Artificial Society and Social Simulation*, **17**, 4. http://jasss.soc.surrey.ac.uk/17/4/reviews/1.html.

*Cowen S.K. (2014), "Secrets and Misperceptions: The Creation of Self-Fulfilling Illusions", *Sociological Science*, **1**: 466–92.

Demeulenaere, P. (2011), "Introduction", in P. Demeulenaere (ed.), *Analytical Sociology and Social Mechanisms*, Cambridge, UK: Cambridge University Press, pp. 1–30.

*DiMaggio P., and F. Garip (2011), "How Network Externalities Can Exacerbate Intergroup Inequality", *American Journal of Sociology*, **116**, 6: 1887–933.

Edling C., and J. Rydgren (2014), "Analytical Sociology: Bringing Culture and Identity Back In", *Sociologica*, **2**. https://doi.org/10.2383/78818.

Edling C., and J. Rydgren (2016), "Social Mechanisms in Empirical Sociology: Introduction to Special Issue", *American Behavioral Scientist*, **60**, 10: 1135–45.

Erikson E. (2013), "Formalist and Relationalist Theory in Social Network Analysis", *Sociological Theory*, **31**, 3: 219–42.

Erikson E. (2017), "Networks and Network Theory: Possible Directions for Unification", in C.E. Benzecry, M. Krause, and I.A. Reed (eds), *Social Theory Now*, Chicago: University of Chicago Press, pp. 278–305.

Ermakoff I. (2017), "On the Frontiers of Rational Choice", in C.E. Benzecry, M. Krause, and I.A. Reed (eds), *Social Theory Now*, Chicago: University of Chicago Press, pp. 162–200.

*Goldberg A., and S.K. Stein (2018), "Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation", *American Sociological Review*, **83**, 5: 893–932.

Gross N. (2009), "A Pragmatist Theory of Social Mechanisms", *American Sociological Review*, **74**: 358–79.

Hedström P. (2005), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge, UK: Cambridge University Press.

Hedström P., and P. Bearman (2009), "What is Analytical Sociology all about?", in P. Bearman and P. Hedström (eds), *The Oxford Handbook of Analytical Sociology*, Oxford, UK: Oxford University Press, Ch. 1.

Hedström P., and G. Manzo (2015), "Recent Trends in Agent-Based Computational Research: A Brief Introduction", *Sociological Methods and Research*, **44**, 2: 179–85.

Hedström P., and R. Swedberg (1996), "Social Mechanisms", *Acta Sociologica*, **39**, 3: 281–308.

Hedström P., and R. Swedberg (eds) (1998a), *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge UK: Cambridge University Press.

Hedström P., and R. Swedberg (1998b) "Social Mechanisms: An introductory Essay", in *Social Mechanisms. An Analytical Approach to Social Theory*, Cambridge UK: Cambridge University Press, Ch. 1.

Hedström, P. and L. Udehn (2009), "Analytical Sociology and Theories of Middle-Range", in P. Bearman and P. Hedström (eds), *The Oxford Handbook of Analytical Sociology*, Oxford, UK: Oxford University Press, Ch. 2.

Gianluca Manzo - 9781789906851

Hedström P., and P. Ylikoski (2014), "Analytical Sociology and Rational-Choice Theory, in G. Manzo (ed.), *Analytical Sociology: Actions and Networks*, Chichester, UK: Wiley, pp. 53–70.
Jaeger M.M., and R. Breen (2016), "A Dynamic Model of Cultural Reproduction", *American Journal of Sociology*, **121**, 4: 1079–1115.
Kalter F., and C. Kroneberg (2014), "Between Mechanism Talk and Mechanism Cult: New Emphases in Explanatory Sociology and Empirical Research", in J. Friedrichs and A. Nonnenmacher (eds), *Kölner Zeitschrift für Soziologie und Sozialpsychologie, Special Issue* **54***: Social Contexts and Social Mechanisms*: 91–115.
Keuschnigg M., N. Lovsjö, and P. Hedström (2018), "Analytical Sociology and Computational Social Science", *Journal of Computational Social Science*. https://doi.org/10.1007/s42001-017-0006-5.
*Keuschnigg M., and T. Wolbring (2015), "Disorder, Social Capital, and Norm Violation: Three Field Experiments on the Broken Windows Thesis", *Rationality and Society*, **27**, 1: 96–126.
Kim, Lanu, C. Adolph, J.D. West, and K. Stovel (2020), "The Influence of Changing Marginals on Measures of Inequality in Scholarly Citations: Evidence of Bias and a Resampling Correction", *Sociological Science*, **7**: 314–41.
*Kroneberg C., and A. Wimmer (2012), "Struggling Over the Boundaries of Belonging. A Formal Model of Nation Building, Ethnic Closure, and Populism", *American Journal of Sociology*, 118, **1**: 176–230.
*Leszczensky L., and S. Pink (2019), "What Drives Ethnic Homophily? A Relational Approach on How Ethnic Identification Moderates Preferences for Same-Ethnic Friends", *American Sociological Review*, **84**, 3: 394–419.
Little D. (2012), "Analytical Sociology and the Rest of Sociology", *Sociologica*, **1**. https://doi.org/10.2383/36894.
Lizardo O. (2012), "Analytical Sociology's Superfluous Revolution: Comment on Little, *Sociologica*, **1**. https://doi.org/10.2383/36902.
Manzo G. (2007), "Variables, Mechanisms, and Simulations: Can the Three Methods Be Synthesized? A Critical Analysis of the Literature, *Revue Française de Sociologie*, **48**, 35–71.
Manzo G. (2010), "Analytical Sociology and its Critics", *European Journal of Sociology*, **51**, 1: 129–70.
Manzo G. (2011), "The Whole Is Greater Than the Sum of Its Parts: Some Remarks on the *Oxford Handbook of Analytical Sociology*", *European Sociological Review*, **27**, 6: 829–35.
Manzo G. (2012a), "Full and Sketched Micro-Foundations: The Odd Resurgence of a Dubious Distinction", *Sociologica*, **1**. https://doi.org/10.2383/36900.
Manzo G. (2012b), "Reason-Based Explanations and Analytical Sociology: A Rejoinder to Boudon", *European Journal of Social Sciences*, **50**, 2: 35–65.
Manzo G. (2013), "Is Rational Choice Theory *Still* a Rational Choice of Theory?", *Social Science Information*, **52**, 3: 361–82.
Manzo G. (2014), "Data, Generative Models, and Mechanisms: More on the Principles of Analytical Sociology", in G. Manzo (ed.), *Analytical Sociology: Actions and Networks*, Chichester, UK: Wiley, pp. 4–52.
Manzo G., and D. Baldassarri (2015), "Heuristics, Interactions, and Status Hierarchies: An Agent-Based Model of Deference Exchange", *Sociological Methods and Research*, **44**, 3:329–87.
Manzo G., S. Gabbriellini, V. Roux, and F.N. M'Mbogori (2018), "Complex Contagions and the Diffusion of Innovations: Evidence From a Small-N Study", *Journal of Archaeological Method and Theory*, **25**, 4: 1109–54.
Marchionni C., and P. Ylikoski (2013), "Generative Explanation and Individualism in Agent-Based Simulation, *Philosophy of the Social Sciences*, **43**, 3: 323–40.
Mayntz R. (2004), "Mechanisms in the Analysis of Macro-Social Phenomena", *Philosophy of the Social Sciences*, **34**, 2: 237–59.
Morgan S.L. (2005), *On the Edge of Commitment: Educational Attainment and Race in the United States*, Stanford, CA: Stanford University Press.
Norkus Z. (2005), "Mechanisms as Miracle Makers? The Rise and the Inconsistence of the 'Mechanismic Approach' in Social Science and History", *History and Theory*, **44**: 348–72.
Opp K.-D. (2005), "Explanations by Mechanisms in the Social Sciences. Problems, Advantages, and Alternatives", *Mind and Society*, **4**, 2: 163–78.
Opp K.-D. (2007), "Peter Hedström: Dissecting the Social. On the Principles of Analytical Sociology", *European Sociological Review*, **23**, 1: 115–22.
Opp K.-D. (2013), "What Is Analytical Sociology? Strengths and Weaknesses of a New Sociological Research Program", *Social Science Information*, **52**, 3: 329–60.
*Polavieja J. (2012), "Socially-Embedded Investments: Explaining Gender Differences in Job-Specific Skills", *American Journal of Sociology*, **118**, 3: 592–634.
Romero M. (2020), "Sociology Engaged in Social Justice", *American Sociological Review*, **85**, 1: 1–30.
Reiss J. (2007), "Do We Need Mechanisms in the Social Sciences", *Philosophy of the Social Sciences*, **37**, 2: 163–84.
Sawyer R. Keith (2007), "Peter Hedström: Dissecting the Social. On the Principles of Analytical Sociology", *Philosophy of the Social Sciences*, **37**: 255–60.
Snijders Tom A.B., and Christian E.G. Steglich (2015), "Representing Micro-Macro Linkages by Actor-Based Dynamic Network Models", *Sociological Methods and Research*, **44**, 2: 222–71.
Sørensen A. (1998), "Theoretical Mechanisms and the Empirical Study of Social Processes", in P. Hedström and

Gianluca Manzo - 9781789906851

R. Swedberg (eds), *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge, UK: Cambridge University Press, pp. 238–66.

Steel D. (2004), "Social Mechanisms and Causal Inference", *Philosophy of the Social Sciences*, **34**, 1: 55–78.

*Skvoretz J. (2013), "Diversity, Integration, and Social Ties: Attraction versus Repulsion as Drivers of Intra- and Intergroup Relations", *American Journal of Sociology*, **119**, 2: 486–517.

Turner J.H. (2019), "The More American Sociology Seeks to Become a Politically-Relevant Discipline, the More Irrelevant It Becomes to Solving Societal Problems", *The American Sociologist*, **50**: 456–87.

*van de Rijt A. (2019), "Self-Correcting Dynamics in Social Influence Processes", *American Journal of Sociology*, **124**, 5: 1468–95.

Van Noorden R., B. Maher, and R. Nuzzo (2014), "The Top 100 Papers", *Nature*, **514**, 7524: 550–3. https://doi.org/10.1038/514550a.

Winship C., and Morgan, S.L. (2014), *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, 2nd edition, Cambridge, UK: Cambridge University Press.

Ylikoski P. (2012), "Micro, Macro, and Mechanisms", in H. Kincaid (ed.), *The Oxford Handbook of Philosophy of Social Science*, Oxford, UK: Oxford University Press, pp. 21–45.

Ylikoski P. (2013), "The (Hopefully) Last Stand of the Covering-Law Theory: A reply to Opp", *Social Science Information*, **52**, 3: 383–93.

Gianluca Manzo - 9781789906851

1. Understanding the Coleman boat

Petri Ylikoski

1. INTRODUCTION

The Coleman diagram is one of the most famous theoretical diagrams in sociology. It is often referred to as *Coleman's boat* and, in the German-speaking world, it is sometimes called *Coleman's bathtub*. In this chapter, I will call it the *diagram*. James Coleman introduced the diagram in his 1986 paper and used it extensively in Coleman (1987) and his *magnum opus, Foundations of Social Theory* (1990). The diagram did not catch on immediately, but especially after Hedström and Swedberg used a modified version of it in their 1998 anthology about social mechanisms, the diagram has become a popular representation of the micro-macro challenge in analytical sociology, and also outside of it.

The diagram is interesting for two reasons. First, it addresses one of the core issues in social theory: how to relate micro and macro. Sociologists are interested in explaining the emergence, persistence and change of large-scale social events, properties and processes. Their relationship to smaller-scale social events and processes (like individual behavior, social interactions etc.) provides a thorny set of problems. The diagram offers a way to think about these micro-macro relationships.

Second, the diagram is a scarce visual representation in sociological theory. Most diagrams by sociological theorists are mere summaries of verbal argument or otherwise primitive compared to the visual representations developed in other sciences (Lynch 1991; Turner 2010; Swedberg 2016). In contrast, the Coleman diagram is a representation that can serve as a cognitive tool for thinking, a diagram for theorizing (Swedberg 2016). It is "a genuine tool of production," as Wright Mills (1959, p. 213) puts it. The diagram provides a systematic scheme for articulating social explanations and their presuppositions. The contents of the scheme could, in principle, be verbally articulated, but the easy to remember diagram provides a more cognitively efficient representation. The abstract form of the diagram makes it easy to adapt to various applications. It raises a sequence of questions that a sociologist must answer to claim to have a theoretical understanding of the social phenomenon.

However, an abstract diagram can be interpreted in many different ways. Even a cursory look at various interpretations, modifications and applications of the diagrams demonstrate that there is no general agreement about what the diagram represents. For example, many critics have taken the diagram as a summary of Coleman's theoretical ideas. Thus, it has been interpreted as a statement of implications of methodological individualism, rational choice theory or, more generally, reductionism. It has also been interpreted as an argument for the causal impotence of macro properties. I will argue that all these interpretations are mistaken. They reduce the diagram to a (bad) theory picture (Swedberg 2016) that loses the productive contribution the diagram can make. In the following I will discuss the diagram as a cognitive tool that is independent of Coleman's other theoretical commitments. Also, people who might disagree with Coleman on methodological individualism or rational choice theory can make productive use of the diagram. It is a tool for thinking, not a metaphysical or theoretical statement.

Petri Ylikoski - 9781789906851

This chapter aims to provide a rational reconstruction of the diagram as a tool for social scientific theorizing. I will show how Coleman uses the diagram and how it can be used as a tool for theoretical thinking. I will also demonstrate how it can be used to clarify the nature of the micro-macro challenge in social explanation. The structure of the chapter is the following. Section 2 will describe the diagram and its elements and how Coleman used the diagram in his work. Section 3 follows the arrows of the diagram and discusses their interpretation and presuppositions. It also introduces an additional arrow to the diagram to make the micro–macro transition more understandable.

2. THE NODES AND ARROWS

This chapter will focus on how Coleman used the diagram and how it can be improved. Later authors have given the diagram many alternative interpretations that I cannot discuss here.[1] To facilitate the discussion, I use capital letters to refer to nodes of the diagram and numbers to the arrows, as shown in Figure 1.1. Coleman himself did not use letters to name the nodes, as in his diagrams the nodes always contained a reference to relevant causal factor or variable. Similarly, the numbering of the arrows is not present in Coleman's original diagrams, but their use has become standard practice. However, the numbering varies between authors. In some versions of the diagram the arrow between nodes A and D is missing. This absence does not change the meaning of the diagram. The reason I have labeled it with a question mark rather than number will soon become apparent.

2.1 The Nodes

To use the diagram, one has to understand what the nodes and arrows mean. The nodes A and D refer to various sorts of macro social facts that figure as *explanantia* and *explananda* in sociological explanations. A facts are extra-individual social factors that might be cited as causes of social phenomena and might influence individuals. D facts are macro-social facts to be explained. These facts can be quite diverse. Among Coleman's sociological *explananda* are revolutions, economic structures, inequalities, panics, election outcomes and disintegration of families. Coleman's notion of the macro is flexible: its scale can vary from a two-person social interaction to organizations and nation states (Coleman 1986, 1987, 1990; Manzo 2020). Also, the A facts can be quite different in their scale. Coleman says that A facts could be

[1] Coleman did not invent the diagram out of nothing and, regrettably, he did not refer to any of his precursors. One clear precursor can be found in David McClelland's book *The Achieving Society* (1961, p. 47). McClelland's diagram has the same recognizable arrows, and he also uses the same example, Weber's thesis in *Protestant Ethic*. It is less clear whether McClelland used the diagram in the same way as Coleman. (It is also noticeable that there are conceptual precursors, like Hernes 1976 and Boudon 1981, that used alternative visual means to represent the same ideas.) However, it seems that all the later published uses of the diagram refer to Coleman (or to authors who refer to Coleman). Without Coleman the diagram might have been completely forgotten. The diagram has also been modified, changed or reinterpreted in many ways by later authors (see, for example: Hedström and Swedberg 1996; Hedström and Swedberg 1998; Abell 2000; Udéhn 2001; Manzo 2007; Abell, Felin and Foss 2008; Sampson 2012; Thornton, Ocasio and Loundsbury 2012) However, apart from Hedström's and Swedberg's 1998 mechanistic interpretation, these modifications have not been adopted for more general use. For the history of the diagram see, Barbera 2007 and Raub and Voss 2017.

Petri Ylikoski - 9781789906851

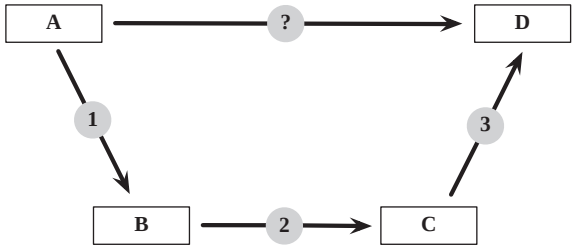


Figure 1.1 The Coleman diagram

"characteristics of their social environments, ranging from family to friends to larger social contexts" (1990, p. 1). The A facts are not privileged: in principle, all A's can (and should) serve as D's for other explanations. In other words, there are no unexplainable social facts. It is crucial to understand that A and D do not refer to the whole of "macro-level"—assuming that it is even a coherent notion—but only to specific facts relevant for the hypothesis under consideration.

The node B refers to properties of agents and their situations. Most of Coleman's examples of B-facts are about values, opportunities or interests, but this set can also be extended to analogical theoretical concepts like beliefs, desires, goals, values, preferences, motives, emotions, habits (or habitus), routines, scripts, heuristics, cognitive schema or identities (social and psychological), as all these notions play a similar role in sociological explanation. They are theoretical concepts used in the explanation of human behavior, and their purpose is to mediate between social influences and individual behavior. Some of these notions might ultimately be too vague to be useful in social science, but the evaluation of these concepts is not a concern of this chapter. The B-facts are important because they explain behavior (C facts), and they mediate the influence of A-facts on behavior. While there might be theoretical reasons to avoid some of these notions, the explanation of action requires that some such theoretical concepts are employed. Sociologists are not primarily interested in explaining B-facts themselves; instead, B-facts are a way to link A-facts to C-facts.

The node C refers to behavioral outcomes. Depending on the context they could be choices, behaviors or actions. These individual outcomes are important *explananda* for sociology, but they are also a step in explaining macro outcomes (D-facts). Coleman allows that B and C facts are not about individual persons. They could "… be institutions within the system or subgroups" (1990, p. 2). They could also be corporate agents. Coleman's notion of micro is flexible, as is his notion of the macro. This flexibility implies that the diagram cannot be understood as an argument for methodological individualism (contra Jepperson and Meyer 2011). For Coleman, the agency is of primary importance, not individuality (or personhood) of agents (Ylikoski 2017). Thus, he writes, "for some investigations, corporate bodies such as formal organizations are usefully regarded as purposive actors, though in other research and theory in sociology, the coherence of their action would itself be taken as problematic" (1986, p. 1312).

2.2 The Arrows

The arrows are more difficult to interpret. They are sometimes misunderstood as representing relationships of reduction or logical derivation. However, the best way to understand them

Petri Ylikoski - 9781789906851

is to take as explanatory relationships that are based on causal dependencies. It is clear that Coleman intends the arrows 1, 2 and 3 to be interpreted causally. For example, he articulates the arrows in Weber example in the following way (1990, p. 8):

- Arrow 1: "Protestant religious doctrine generates certain values in its adherents."
- Arrow 2: "Individuals with certain values adopt certain kinds of orientations to economic behavior."
- Arrow 3: "Certain orientations to economic behavior on the part of individuals help bring about capitalist economics organization in a society."

All of these are straightforward causal claims. "Generation" and "bringing about" are everyday causal notions referring to causal processes. It is also evident that in the second sentence the values are assumed to influence the economic activities causally.

For current purposes, it is sufficient to accept that the arrows reflect explanatory dependencies. The more detailed interpretation of these dependencies is contingent on one's account explanation.[2] My simple suggestion is that the arrows are interpreted in terms of difference-making. Thus, for example, in the case of Arrow 1 some change in A-variable brings about a change in B-variable. This is a kind of minimal explanatory dependency that is easy to conceive in terms of (imagined) causal interventions: changing A to A* would bring about change in B to B*. For example, parliament changes the legislation about parental leaves (the intervention on A-variable), which brings about changes in workers' opportunities to take care of their small children (the change in B-variable), which in turn brings about changes in their annual working time (C-variable). This idea is fully compatible with mechanisms-based thinking: one can always ask by which mechanisms these changes were brought about. However, one might not have that information available, or there are multiple causal pathways (mechanisms) at work that might in opposition to each other, so one would still like to know whether there is some sort of net effect on B. In addition, the difference-making interpretation provides an excellent way to conceptualize structural presuppositions of macro-micro and micro-macro claims: they are background conditions—or a causal field (Mackie 1974)—that make the foreground causal claims possible.

It is also useful to think about the arrows in terms of questions they answer. Arrow 1 is associated with the question: How does the macro change affect the agents? Arrow 2 with the question: How does the behavior of agents change? and Arrow 3 with the question: How do changes in behavior bring about macro consequences? The questions can be formulated in various ways, but formulating them in terms of how-questions highlights the interest in the mechanisms of change: it is not enough to know that certain changes are causally related, we also wish to know how the changes happen.

Finally, there is the arrow at the top that I have labeled with a question mark. This arrow is sometimes absent from Coleman's own diagrams.[3] It is the starting point of the analysis that

---

[2] Coleman himself did not explicitly advocate any specific account of explanation. The Dutch interpretation that relies on the covering-law account of explanation and reconstructs the diagram in terms of deductive arguments (see Lindenberg 1981; Raub, Buskens and van Assen 2011 for details). The alternative Swedish interpretation (Hedström and Swedberg 1998), interprets the arrows in terms of causal mechanisms.

[3] Coleman had the arrow in the diagram in his 1987 and 1990 publications, but not in his 1986 and 1992. There is no reason to suppose that this reflects any changes in his thinking.

Petri Ylikoski - 9781789906851

employs the diagram. It can be an observed or hypothetical connection between two macro variables. In such cases the purpose of the diagram is to help to determine whether the relationship between A and D is genuinely causal and, if it is, why it holds. However, D might also be unknown; in such cases, the goal is to reason through the diagram and figure out *what could be* the macro consequences of a change in A.

2.3 The Motivations

The motivation to use the diagram is to make macro causal claims more secure by finding the underlying causal mechanisms. For example if the university leadership believes that changes in professors' salaries increase the scientific quality of research produced by the university as a whole, it is a good idea to articulate how the supposed causal chains are assumed to work. If one cannot provide a plausible and empirically supported story, there is every reason to be skeptical about the claimed causal dependence. In addition to *the justification of causal claims*, the other added value provided by the diagram is *theoretical understanding*. Rather than just citing a relationship between two macro variables, the ideal sociological explanation provides an account for how the suggested cause brings about a change in individuals' (or other agents') beliefs, desires or opportunities; how these changes, in turn, bring about changes in their behavior and, finally, shows how these behavioral changes bring about the macro change to be explained. In other words it provides an action-based explanation. The diagram shows how such explanations have three analytically distinct explanatory subtasks. For this reason, the diagram is based on the rationale of mechanism-based explanation. The information about the mechanisms helps to understand under which background conditions the causal dependency holds, how it would change when some of the background conditions change and what can be done to prevent or sustain the causal outcome (Ylikoski 2011).

While it cannot be expected that a single study or publication could provide the complete story, the diagram tells us what parts a full sociological explanation would have to have. It provides a skeleton for an ideal explanatory narrative. For this reason it can be utilized as a tool for thinking about particular sociological explanations. Coleman offers many examples of this. First, he uses the diagram to *illustrate* specific explanatory ideas—for example the marriage and other matching markets (1987, pp. 159–160) and problems with collective decision making (1987, pp. 169–71; 1990, pp. 400–2)—and to *compare* competing explanations, for example the explanations for panic in a burning theatre (1987, pp. 161–3; see also 1990, pp. 203–9). More importantly, he employs the diagram to *point out missing crucial elements* in the proposed explanatory narratives. This is the main point of his discussion of Weber's thesis in *Protestant Ethic* (1987, pp. 154–5; 1990, pp. 6–10). He argues that

> … some sort of combined or joint or aggregate effect of the economic behavior of many individuals in bringing about capitalist development is being proposed. It is here, however, that Weber's analysis is almost totally silent. (1990, p. 9)

Finally, Coleman uses the diagram to *articulate structural assumptions of explanations*. Good examples of this are his discussions of the influence of the educational inequality on income inequality (1987, pp. 163–6) and the impact of training programs on youth employment (1987, pp. 166–8; 1990, pp. 642–5). In both cases the diagram is used to trigger articulations of presuppositions of suggested explanations so that their credibility can be evaluated.

Petri Ylikoski - 9781789906851

The diagram can also be used in thinking about foundational issues in social theory. First, it is useful in highlighting the non-trivial nature of the problem of micro-foundations. Both explanations of C-facts and D-facts involve structural assumptions that are crucial for understanding the explanatory dependencies. Here is Coleman's key message: the micro-macro relationships are aggregative only in exceptional cases. Second, the diagram can be used to illustrate the role of the "theory of action" in a sociological explanation. While pure BC-explanations fall mainly outside the scope of sociology, both ABC and BCD explanations presuppose BC-processes, which implies that assumptions about psychological processes play an important role in sociological explanations. Finally, the diagram also presents the basic idea of mechanism-based explanation already explained above: sociological explanations that merely connect two macro variables (direct AD-explanations) are not theoretically satisfactory, and they need to be supplemented by an account of the micro process underlying it. The point of mechanism-based explanation is to show how the causal (or non-causal) relationship between A and D is brought about. This helps us to both justify or criticize the suggested causal connection and have an explanatory understanding of it.

3. THINKING THROUGH THE DIAGRAM

In this section, I will take a closer look at the elements of the diagram and discuss some difficulties related to their interpretation. I argue that while the interpretation of the diagram is more complicated than usually assumed, the ambiguities and problems do not pose a fundamental challenge for the usefulness of the diagram in analyzing sociological explanations.

3.1 Macro to Micro

Causal influences from macro to micro are a central concern in sociology. One could say that they are a defining feature of a distinctly sociological approach to the social world. This is apparent when we consider the alternative diagram shown in Figure 1.2. This diagram would be a natural way to represent a vision of social science—asociological atomism—that sociologists often contrast themselves with. An atomist begins with ahistorical assumptions about human nature (as in J.S. Mill's idea of social science), or with purely stipulated preferences or tastes of individuals (as in sociologists' stereotype of mainstream economics), and proceeds to explain social outcomes. In contrast, sociology is interested in social factors influencing individuals and their behavior. This does not imply that sociologists would assume that all influences are social, but it is typical for them to assume that social influences (during the life course of an individual) are of crucial importance. One of Coleman's motivations for introducing the diagram was to show that sociologists have been too occupied with macro-to-micro influences at the cost of attending to micro-to-macro influences.[4]

As suggested above, the macro social facts—the A-facts—should be considered broadly. In Coleman's own work, they are quite diverse. For example, there are such things as corporate action or social policy (1990, p. 646), legislation creating a job-training program (1987, p. 167;

[4] However, Coleman (1993) later commented, "Were I writing the book over again, I would give considerably more attention to this macro-micro relation" (p. 63). In other words, he thought that both are important.

Petri Ylikoski - 9781789906851

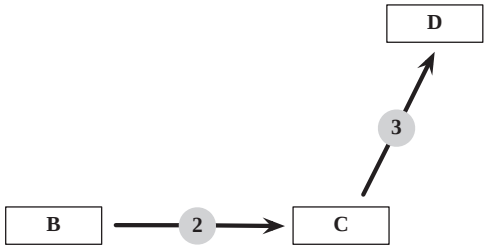


Figure 1.2 Atomistic version of the diagram

1990, p. 643), improved social conditions (1987, p. 156; 1990, pp. 10, 478), dissatisfaction (1986, p. 1321), Protestant religious doctrine (1986, p. 1322; 1987, p. 155; 1990, p. 8), change in birth rate (1987, p. 160), a collective fear created by the sound of fire alarm in a crowded theatre (1986, p. 1321; 1987, pp. 161–2), educational level of society (1987, pp. 165–6), the set of alternatives for collective decisions (1987, p.169), the means of production (in the Marxist theory) (1986, p. 23), and distribution of resources (1992, p. 121). What is common to these diverse examples is that they all refer to social facts that are "larger" than individuals. While it is impossible to see all of them as belonging to some unique or comprehensive "social level," their shared characteristics are apparent when we think of the micro-macro contrast in terms of scale (Ylikoski 2012, 2014). In this view macro facts can be of different "sizes": from local social interactions to whole social systems; thus, there is no need to assume that there is some fixed macro (or meso) level.

An advantage of thinking macro in terms of scale is that it makes it unnecessary to conceive macro-to-micro influences in terms of "downward causation." This metaphorical notion, laden with strong metaphysical assumptions, can be replaced with the down-to-earth idea of large-scale social facts influencing smaller-scale facts about individuals (or other agents). In terms of difference-making, it is a question of how changes in large-scale facts bring about changes in smaller-scale facts. Furthermore, it is easy to see that dependencies like this are contingent on relevant (structural, institutional, cultural etc.) background conditions: if the background conditions were different, the dependence would be different or absent.

However, there is a complication related to explanatory dependencies between A and B facts. While Coleman seems to think that these dependencies are straightforwardly causal, this cannot be the case. Consider Coleman's favorite example, Max Weber's Protestant ethic thesis. In Coleman's reconstruction, we have the Protestant religious doctrine (the A-fact) influencing the values of individuals (the B-fact).[5] In the case of a person we could say that replacing the Catholic religious doctrine with the Protestant doctrine causally changes a person's values. The doctrinal change is first, or at least it starts first, and the changes in values follow. However, in Coleman's analysis, the adoption of the Protestant doctrine is assumed to be a macro fact. The idea is that when the relevant groups adopted Protestantism, their members' values changed. While there may be some value changes that are really causal consequences of the conversion to Protestantism, it seems that these are not changes that are the main focus of Weber's analysis. At least some of the relevant values have a more intimate relationship to doctrines

[5] Weber did not talk about "values" in the sense Coleman uses the term. However I will follow Coleman's terminology to keep the example simple.

Petri Ylikoski - 9781789906851

of the Protestant sects. (The religious doctrines were an important source of evidence about these values for Weber.) Thus, one could say that the group's adoption of the religious doctrine partly *consists* of its members adopting specific new values (Ylikoski 2013). While there is more to the Protestant doctrine than these values, they are part of it. Thus, one can say that as a consequence of the group's adoption of the Protestant doctrine, its members' values changed, but one cannot say that the adoption of the doctrine caused the change in values.

It might be possible to fiddle with these concepts in a way that would turn this example causal. After all, Coleman's discussion of the case is so sketchy that there is a lot of room for alternative interpretations. However, the problem is more general. Consider how the demographic change caused by war affects the number of potential marriage partners available for women, how the law allowing same-sex marriage changes the opportunities of same-sex couples to arrange their legal relationship or how improved the educational level of society is related to the education of individuals. In all these cases, it is plausible to say that the cited A-facts partially consist of the mentioned B-facts. However, at the same time, some of the important consequences might well be causal. Thus, it seems that A-B explanatory dependencies are based on (various) mixtures of causal and constitutive relationships (Ylikoski 2013).

Let us look at this in more detail. Consider two processes:

- a macro process A-A* (the adoption of the Protestant doctrine within the community)
- a micro process B-B* (the changes in the values (or life-orientation) of the individual community members)

What are the relationships between these two processes? First, we can say that the A-A* process *involves* the B-B* process. However, it involves more, so we cannot simply say that the B-B* process constitutes the A-A* process. Second, we can say that bringing about A-A* would also bring about B-B*. We could even say that the process A-A* would bring about the end-state B*. Third, we cannot say that A-A* is the cause of B-B*, nor can we say that the latter is logically implied by the former. Finally, B* may be later than A*. This would provide evidence there are real causal processes involved in the chain of changes. Based on these observations we can conclude that the A-B* relationship works pretty much like a causal relationship, although it consists of a mix of causal and constitutive relationships. Thus, this complication does not challenge the intelligibility or coherence of the thinking underlying the diagram. This conclusion is further supported when we notice that the real *explananda* in the diagram are C-facts and D-facts. Whatever is the relationship between A- and B-facts, the link of A-facts to C-facts is causal.

Another complication are the things that are not represented in the diagram. All the arrows in the diagram abstract away from causal details: they have background conditions, and they might consist of multiple causal pathways. Let us first discuss the possibility of multiple pathways. For example, A-facts might influence the relevant B-facts in more than one way. It is also possible that some of these causal pathways have opposing influences. For this reason, Arrow 1 basically represents a net causal effect. The multiplicity of causal pathways could, in principle, be represented by multiple parallel arrows. However, this would, apart from demanding more from artistic skill, make the diagram much more complicated without clear cognitive advantage. The diagram would be more challenging to read and memorize. The reader just has to understand the abstraction level of the diagram and recognize that there might be multiple pathways.

Petri Ylikoski - 9781789906851

Presuppositions of the dependencies represented by the arrow are the second thing that the diagram does not show. As suggested, the dependencies hold against the background of the causal field. What this causal field consists of is often a significant challenge for a theorist, and it poses problems for theoretical generalization as many of the presuppositions are structural. As with the case of multiple pathways the reader has to recognize the role of causal background conditions without them being represented. As long as one remembers the rather obvious fact that macro-to-micro and micro-to-macro influences must have structural presuppositions there is no reason to complicate the diagram by adding multiple new elements.

3.2 Theory of Action

The arrow between B and C captures the role of theory of action in sociology. It bridges between B-states and C-outcomes. As already noted Coleman has quite a liberal notion of agency. He is quite ready to accept non-individual agents, at least for some explanatory purposes (1986, p. 1312; 1990, pp. 2–5). This is not an isolated feature of his thinking, for one of his major sociological claims is that various forms of corporate agency (corporations, states, unions, parties etc.) have significantly changed our societies over the last two centuries. Along with many other methodological individualists (Ylikoski 2017), Coleman is more concerned with the agency than with individuality itself. However, as Coleman does not provide a discussion of non-individual agency in the context of the diagram, I will not discuss it further in this chapter.

Another important point is related to the Coleman's view of the division of labor between social sciences and psychology. He says:

> … a natural stopping point for the social sciences (although not psychology) is the level of the individual—and that, although an explanation which explains the behavior of a social system by the actions and orientations of some entities between the system level and the individual level may be adequate for the purpose at hand, a more fundamental explanation based on the actions and orientations of individuals is more generally satisfactory. (1990, p. 4)

Coleman expresses here the idea that social scientific explanations bottom out at the level of individual action. The idea is that social scientists can take it for granted that there are intentional agents. However, this is not to be treated as an ultimately unexplainable fact. However, it is the task of psychology (and neurosciences) to explain how we can have these capabilities.

This division of scientific labor implies that social scientists do not have to focus on all intricate details of individual agency. The task of sociology is to understand complex larger-scale social facts, so it should operate with relatively simple schemes of action explanation. Coleman identifies an important trade-off:

> … it is especially important that the individual-action component remain simple […] a trade-off between complexity in the other two components and complexity in this component must be made if the overall theory is to remain manageable. (Coleman 1990, p. 19)

His message is that the sociological theory of action can be relatively shallow from the point of view of psychology, as the explanation of individual cognition is not the task of the social sciences. What is needed is a scheme that allows bridging A-facts to C-outcomes. This perspective helps to understand why Coleman adopted the rational choice theory as his preferred theory of action. The rational choice theory provides a very simple version of folk

Petri Ylikoski - 9781789906851

psychology that can be adapted for mathematical modeling. While the rational choice model misses many facts about individual psychology, gets other facts wrong, and is all too simple, it is, in Coleman's judgment, a useful workhorse for sociological theorizing. This assessment can be challenged. There is no reason to assume that sociologists have to use the same theory of action in all contexts. They can adopt a more pragmatic attitude and give up the idea of foundational sociological theory of action (Hedström and Ylikoski 2014).

Another advantage of the more pragmatic approach is that it is also more naturalistic. Coleman (1986) is attached to the "humanistically congenial image of man" provided by the model of purposive action. He believed that it would provide an interface between sociology, everyday reasoning, economics and law. In light of more recent developments in psychology and cognitive sciences, it is clear that Coleman was too optimistic about the prospects of a folk psychology–based scheme. We may have to give up Coleman's rather strict division of labor between psychology and social sciences. For many purposes, explanations that appeal to sub-personal cognitive processes might be better than those based on everyday folk psychology. The incorporation of insights from cognitive sciences, when relevant, does not make the diagram irrelevant. Instead, it highlights its importance, as it helps sociologists to keep their eye on real issues, e.g. the macro facts to be explained and the micro-macro transitions, rather than the foundational theory of action. Thus, while Coleman might have been attached to some form of intentional fundamentalism (Ylikoski 2012), this is not an essential part of the diagram. The diagram is based on the principles of mechanism-based explanation in the social sciences, and when we replace the foundational theory of action with psychological theories based on cognitive mechanisms, it reflects the ideals of mechanism-based explanation even more consistently.

The third element that the diagram does not represent consists of the heterogeneity of agents. The agents represented in the diagram might have different initial beliefs, goals, cognitive schema, routines and identities. They also hold different social positions and have different resources and opportunities. There is no reason to assume that the same A fact would influence them in precisely the same way. Thus, their behaviors (the C facts) would also be different. The heterogeneity of agents is an important social fact, and it is often of crucial explanatory importance. Modeling of this heterogeneity is intricate and social scientists have often tried to solve the problem by employing concepts like the representative agent (Hartley 1997). Only recently have agent-based simulation tools made it possible to model how the heterogeneity of agents influences the macro outcomes (Manzo 2020). Nothing in the logic of the diagram implies that we should think that intellectual compromises like representative agents could solve the problem of micro-foundations. Like background conditions and multiple pathways the heterogeneity of agents is a fundamental fact about the social world that should be remembered when using the diagram.

3.3 Micro to Macro

For Coleman, the third arrow is the most crucial part of the diagram. The main reason he started to utilize the diagram was to highlight the challenges posed by the micro-macro transition. According to him, both in sociology and economics, not enough attention is given to the complexities of the micro-macro relationship. It is common to assume that the macro order is just an aggregate effect of micro facts about individuals. Notice that this is not just a problem for individualistic theories. While holistic theories tend to highlight contextuality and

Petri Ylikoski - 9781789906851

complexity, in practice their micro-macro assumptions are rather simplistic. It is quite often assumed that macro facts simply reflect the relevant micro facts and vice versa. Schelling's (1978) famous checkerboard model shows how this assumption is wrong in even very simple settings. Although the residents do not favour segregation, they might still end up in a highly segregated neighbourhood. Similarly, it cannot be inferred from the segregation of the area that the residents favour this state of affairs. Similarly, Granovetter's (1978) threshold models of collective action demonstrate that it does not make sense to assume that the collective activities reflect the average attitudes of participating individuals. In fact, the averages can be highly uninformative about macro facts. Even these elementary examples show that there is a discontinuity between micro and macro.

To make sense of these relationships, Coleman introduced the metaphorical notion of "rules of the game." He suggests that the structural arrangements that determine the micro-macro relationships:

> … can be conceived as the rules of the game, rules which transmit consequences of an individual's action to other individuals and rules which derive macro-outcomes from combinations of individuals' actions. (1990, p. 19)

Coleman did not provide a general definition of the idea, but gave some examples such arrangements as the market and other matching processes. However, he did not spell out the "rules" of these examples. Furthermore, the metaphor is not very helpful in conceiving the missing elements in Weber's historical scenario. The crucial aspects of the micro-macro link are not captured with the idea of rules of the game. This leaves an opportunity for others to improve his work.

The arrows of the diagram give it a dynamical nature, as the observer follows the arrows, s/he also follows the explanatory narrative consisting of the chain of arrows. However, this representation has the disadvantage of being linear and giving an impression of the one-shot process. Sociologically interesting processes are rarely simple one-shot causal chains that begin with a change in the *explanans* variable and then end with a change in the *explanandum*. In most interesting cases the B-facts are not merely influenced by changes in A-facts, but also by the consequences of earlier C-facts.

This is the fourth missing element that is not captured by Coleman's original diagram. The idea is the following: the initial A-facts change B-facts about the agents, which in turn influences their behavior (the C-facts). The feedback loop consists of these behaviors, in turn influencing the attributes and opportunities of the agents. These changes, in turn, influence the agents' behavior, and the loop takes another round.[6] This feedback can be represented by adding an extra arrow to the diagram.[7] Now the unfolding social process can be analyzed as a looping process, and D is the outcome of this process (Figure 1.3).

The endpoint of Arrow 3 is purposefully ambiguous. The behavioral consequences could either affect other agents directly or via the institutional and structural conditions. This is an empirical question; the purpose of the arrow is to remind the researcher that it is important to

---

[6] The idea of a loop is not entirely new. Some authors (Abell 2000; Manzo 2007; Abell, Felin and Foss 2008) have attempted to represent the loop by a chain of boats (where each D is followed by A of the next boat). However, this does not fully solve the problem because the loops do not necessarily involve A and D facts. Thus, chaining boats do not capture those feedback-loops that do not go through A and D.

[7] To keep the order of the arrows logical, I have changed the number of the micro-to-macro arrow.

Petri Ylikoski - 9781789906851

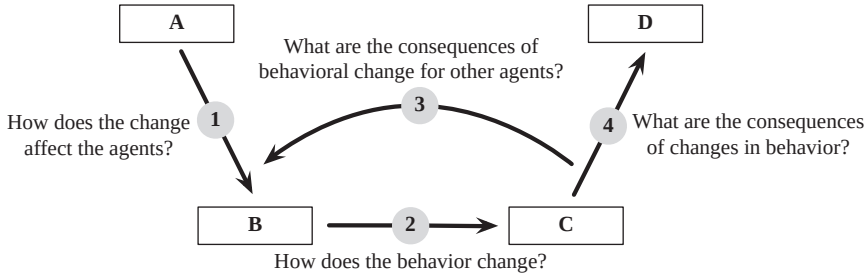


Figure 1.3 Coleman diagram with a feedback loop

look for these feedback loops as they are often crucial to the process of social change. The looping can occur at multiple scales. First, there can be a process by which individual agents learn about their environment, opportunities and other agents. They may find their new opportunities only after some time, or they find them appealing after some delay. Second, institutional or structural features that belong to the background causal field can change.

Schelling's (1978) segregation model helps to illustrate this idea. We can interpret the initial setting as a macro change that makes it possible for families to move freely to new neighborhoods. (The randomness of the initial distribution of agents can be interpreted as randomness with respect to agents' preferences. We could imagine that people were distributed to their houses by some administrative policy.) After this initial macro change, the dissatisfied agents are free to move. Their movements, however, change the composition of both the neighborhood that they are leaving and the neighborhood where they are arriving. This has the consequence that some satisfied agents are now facing an unsatisfactory situation, and maybe some unsatisfied have become satisfied. This is the idea of the feedback loop. The feedback is based on the fundamental fact of *interdependence* between the agents: what an agent does affects what others can do and what they want to do.[8] Notice that the *explanandum* (D-node) is only the final outcome of this process, so producing it could require multiple rounds of feedback.

However, the Schelling example does not fully capture the potential of the diagram. In many interesting cases the result of the process is a *structural change*, e.g. the rules of the game change (Hernes 1976). Consider the effects of *US News* law school rankings studied by Espeland and Sauder (2016). The introduction of the ranking can be regarded as the initial macro change, as it was a large-scale change affecting all relevant agents (Ylikoski 2012). According to Espeland and Sauder, the ranking provided a cognitive map for applicants who were interested in comparing law schools (Arrow 1). With the help of it the earlier very complicated choice was simplified dramatically: it was not possible to compare directly all schools to each other without presupposing any expertise about legal education. Thus, the wide availability and the easy use of the rankings made them a very influential factor in choosing a law school (Arrow 2).

[8] It might be useful to distinguish between two forms of interdependence. *Strategic interdependence* is familiar from the game theory. It deals with the interdependence of choices: the consequences (and rationality) of choice are dependent on what others choose. The feedback loop describes *causal interdependence*: the agent's behaviors influence other agents' later beliefs, desires or opportunities (or other B-facts).

Petri Ylikoski - 9781789906851

To capture the unintended macro consequences of the rankings, we have to focus on the *interdependence of the agents*; in other words, the impact the changes in the applicants' behavior had on other agents. This interdependence is captured by the third arrow representing the causal loops by which the agents influence each other through their behavior. Thus, while most deans initially regarded the *US News* rankings as both incompetent and harmful, their adoption by the applicants made the rankings relevant to them. When the deans started to worry about the rankings, both the content and evaluation criteria of the work by people working in admission and career changed. Similarly, once applicants began to make decisions based on rankings, the employers could start using the rankings as a source of information about the quality of job applicants. Furthermore, when the employers began to use the rankings (or were widely believed to do so), the school rankings became an increasingly important part of the market value of the student's degree. This, in turn, increased the pressure felt by the deans and created a dynamic in which the importance of the rankings was amplified as an increasing number of groups started to react to them.

The created dynamic is crucial for the long-standing changes in legal education produced by the rankings (Arrow 4). The attempts to improve the school's standing involved, among other things, changes in how students were admitted, how the schools utilized their resources, to whom the schools targeted their marketing, and how the career services advised fresh graduates. These changes, in turn, generated structural changes in US legal education; for example, the transformation of the legal curriculum changed the composition of the student body, and the increasing similarity and status competition between the schools (Espeland and Sauder 2016, Chapter 7).

With the added arrow the diagram gets more dynamical as it no longer gives an impression of a one-shot process. With the feedback arrow in place it is much harder to assume that the micro-macro transition is just a simple process of aggregation. It also highlights how thinking with the diagram is a natural way of approaching the complexity of social processes: heterogeneity and interdependence of agents, sensitivity to the background conditions, and various sorts of feedback processes. While only the last one is "officially" represented in the diagram, the questions about others are its natural corollaries.

4. CONCLUSION

This chapter has been an extensive commentary on the Coleman diagram. I have argued that, in contrast to many other sociological diagrams, it is not a mere summary of verbal argument. It is a tool for thinking about one of the central challenges for sociological theory. Coleman's discussion of the diagram was very brief, so I have looked at how he uses the diagram and how these uses can be extended to new cases. The interpretation of the diagram is laden with deep questions about the nature of social theory. As any diagram, it can be given multiple interpretations, and any user of the diagram can use it anyway they find it illuminating.

In my rational reconstruction I have tried to preserve the central idea that the diagram is a tool for thinking. I believe that this was Coleman's original intention, which is most visible in his 1987 chapter. My reconstruction also highlights four important issues for social theorizing: sensitivity to the background conditions, heterogeneity, interdependence of agents and various sorts of feedback loops. To capture the last, I introduce one significant change to the

Petri Ylikoski - 9781789906851

diagram—the feedback arrow—because the addition improves the usability of the diagram and does not involve a radical departure from Coleman's thinking.

REFERENCES

Abell, Peter, (2000), "Sociological theory and rational choice theory", in Turner (ed.), *The Blackwell Companion to Social Theory,* 2nd ed., Oxford: Blackwell: 223–44.

Abell, Peter, Teppo Felin and Nicolas Foss (2008), "Building micro-foundations for the routines, capabilities, and performance Links", *Managerial and Decision Economics,* **29** (6), 489–502.

Barbera, Filippo (2007), "A star is born? The authors, principles and objectives of analytical sociology", *Papers: Revista De Sociologia,* **80**, 31–50.

Boudon, Raymond (1981), *The Logic of Social Action. An Introduction to Sociological Analysis*, London: Routledge and Kegan Paul.

Coleman, James S. (1986), "Social theory, social research, and a theory of action", *American Journal of Sociology*, **91**, 1309–35.

Coleman, James S. (1987), "Microfoundations and macrosocial behavior", in Alexander, Giesen, Münch and Smelser (eds), *The Micro-Macro Link*, Berkeley: University of California Press: 153–73.

Coleman, James S. (1990), *Foundations of Social Theory*, Cambridge, MA: The Belknap Press.

Coleman, James S. (1992), "The vision of foundations of social theory", *Analyse and Kritik,* **14**, 117–28.

Coleman, James S. (1993), "Reply to Blau, Tuomela, Diekman and Baurmann", *Analyse and Kritik,* **15**, 62–9.

Espeland, W.S. and Sauder, M. (2016), *Engines of Anxiety Academic Rankings, Reputation, and Accountability*, New York: Russell Sage Foundation.

Granovetter, M. (1978), "Threshold models of collective behavior", *American Journal of Sociology,* **83**, 1420–43.

Hartley, James (1997), *The Representative Agent in Macroeconomics*, London: Routledge.

Hedström, Peter and Richard Swedberg (1996), "Rational choice, empirical research, and the sociological tradition", *European Sociological Review*, **12**, 127–46.

Hedström, Peter and Richard Swedberg (eds) (1998), *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge: Cambridge University Press.

Hedström, Peter and Petri Ylikoski (2014), "Analytical Sociology and Rational Choice Theory", in Gianluca Manzo (ed.), *Analytical Sociology: Norms, Actions and Networks*, Chichester: Wiley: 57–70.

Hernes, Gudmund (1976), "Structural change in social processes", *American Journal of Sociology*, **82** (3), 513–47.

Jepperson, Ronald and John W. Meyer (2011), "Multiple levels of analysis and the limitations of methodological individualisms", *Sociological Theory,* **29**, 54–73.

Lindenberg, Siegwart (1981), "Erklärung als Modellbau", in W. Schulte (ed.), *Soziologie in der Gesellschaft*, Bremen, Germany: Zentraldruckerei der Universität: 20–35.

Lynch, M. (1991), "Pictures of nothing? Visual construals in social theory", *Sociological Theory,* **9**, 1–21.

Mackie, J.L. (1974), *The Cement of the Universe*, Oxford: Clarendon Press.

Manzo, Gianluca (2007), "Progrès et 'urgence' de la modélisation en sociologie. du concept de 'modèle générateur' et de sa mise en oeuvre", *L'Année sociologique*, **57**, 13–61.

Manzo, Gianluca (2020), "Agent-based models and methodological individualism: are they fundamentally linked?" *L'Année sociologique*, **70**, 197–229.

McClelland, David (1961), *The Achieving Society*, Princeton, NJ: Van Nostrand Co.

Mills, C. Wright (1959), *The Sociological Imagination*, New York: Oxford University Press.

Sampson, Robert (2012), *Great American City: Chicago and the Enduring Neighborhood Effect*, University of Chicago Press.

Raub, Werner, and Thomas Voss (2017), "Micro-macro models in sociology: antecedents of Coleman's diagram", in Benn Jann and Wojtek Przepiorka (eds), *Social Dilemmas, Institutions, and the Evolution of Cooperation*, Berlin: De Gruyter: 11–36.

Raub, Werner, Vincent Buskens and Marcel A.L.M. van Assen (2011), "Micro-macro links and microfoundations in sociology", *The Journal of Mathematical Sociology,* **35**, 1–25.

Schelling, Thomas C. (1978), *Micromotives and Macrobehavior*, London: W. W. Norton.

Swedberg, R. (2016), "Can you visualize theory? On the use of visual thinking in theory pictures, theorizing diagrams, and visual sketches", *Sociological Theory,* **34** (3), 250–75.

Thorton Patricia H., William Ocasio and Michael Lounsbury (2012), *The Institutional Logics Perspective. A New Approach to Culture, Structure, and Process*, Oxford: Oxford University Press.

Turner, Charles (2010), *Investigating Sociological Theory*, London, UK: SAGE.

Udéhn, Lars (2001), *Methodological Individualism: Background, History, and Meaning*, London: Routledge.

Petri Ylikoski - 9781789906851

Ylikoski, Petri (2011), "Social mechanisms and explanatory relevance", in P. Demeulenaere (ed.), *Analytical Sociology and Social Mechanisms*, Cambridge University Press: 154–72.
Ylikoski, Petri (2012), "Micro, macro, and mechanisms", in H. Kincaid (ed.), *The Oxford Handbook of Philosophy of the Social Sciences*, Oxford University Press: 21–45.
Ylikoski, Petri (2013), "Causal and constitutive explanation compared", *Erkenntnis*, **78**, 277–97.
Ylikoski, Petri (2014), "Rethinking micro-macro relationships", in Finn Collin and Julie Zahle (eds), *Rethinking the Individualism-Holism Debate. Essays in Philosophy of Social Science*, Dordrecht: Springer: 117–35.
Ylikoski, Petri (2017), "Methodological individualism", in Lee McIntyre and Alexander Rosenberg (eds), *Routledge Companion to Philosophy of Social Science*, New York: Routledge.

Petri Ylikoski - 9781789906851

2. Analytical sociology, analytic philosophy and analytical Marxism

Daniel Little

Analytical sociology has by now made a very noticeable impact on sociology, especially in the areas of methodology and explanation. The dominant strand of contemporary anglophone philosophy is usually referred to as "analytic philosophy". This encompasses a century-long tradition, including logicism, ordinary language philosophy, philosophy of science, and meta-ethics. Analytic philosophers were heterogeneous in their interests but shared a commitment to the values of conceptual clarity, argumentative rigor and analytical precision. Analytical sociology is a younger intellectual movement, but it too is motivated by an insistence on clarity and analysis of problems into their components. It is worthwhile, therefore, to ask whether there is a relationship between these two "analytical" movements. Two recent histories of analytic philosophy provide a very good exposure to the development of this field since the early twentieth century (Schwartz 2012; Beaney 2017). See also Gellner 1979 and Agassi and Jarvie 2019 for more critical treatments of the tradition.

Another interesting affinity for analytical sociology is in the field of analytical Marxism, an interdisciplinary and highly fruitful area of research in the 1970s through the 1990s. Contributors to the field of analytical Marxism derived largely from the Anglo-American tradition of philosophy and social science, and their goal was to make sense of Marx in terms that largely dispensed with the Hegelian antecedents of some of Marx's ideas. G.A. Cohen (1978) offered an interpretation of historical materialism that made the theory intelligible to analytic philosophers and social scientists; John Roemer proposed to reconstruct Marx's theory of exploitation without making use of the labor theory of value (Roemer 1982); and Jon Elster (1985) and Daniel Little (1986) proposed to understand Marx's key economic and social theories without grounding them in Hegel's philosophy of dialectics. It is very significant that analytical Marxism highlighted some of the ideas and methodological commitments that eventually shaped the field of analytical sociology, including an interest in discovering causal mechanisms, a view of social explanation that gives substantial importance to individual rationality and the need for microfoundations for large-scale causal and structural claims. In fact Diego Gambetta suggests that the term "analytical sociology" was chosen in deliberate analogy with analytical Marxism (quoted from personal correspondence in Manzo 2010, p. 138). So it is worth exploring the connections among all these "analytical" disciplines.

There are several important questions to pose here: did key ideas or methods developed within analytic philosophy play a role in the development of analytical sociology? Does analytic philosophy have anything to offer the field of analytical sociology today? Do analytical sociologists have ideas and theories that can contribute to current topics in analytic philosophy? And most crucially, do analytical sociology, analytic philosophy, and analytical Marxism find common ground in contributing to innovative and insightful ways of better understanding the workings of the social world?

Daniel Little - 9781789906851

Other chapters in this volume provide a great deal of detail about the field of analytical sociology. Here is a brief and abstract description that Peter Hedström provides in *Dissecting the Social*:

> This approach to sociological theorizing and research, which I refer to as 'analytical sociology', seeks to explain complex social processes by carefully dissecting them and then bringing into focus their most important constituent components. The approach focuses on traditional sociological concerns but uses explanatory strategies more often bound in analytical philosophy and behavioural economics. It is an approach that seeks precise, abstract, realistic and action-based explanations for various social phenomena. (Hedström 2005, p. 1)

Dissecting the Social serves as a manifesto for the field of analytical sociology in which Hedström expresses a small handful of intellectual commitments that characterize the field. First is an overriding commitment to provide a standard of clarity and rigor to sociological explanations that the founders of analytical sociology often find lacking in the tradition. Founding contributors such as Peter Hedström, Petri Ylikoski, and Gianluca Manzo prefer to offer compelling explanations of concrete social phenomena rather than large sociological theories intended to encapsulate the nature of the modern world. Analytical sociologists pursue a strategy of decomposition (dissection) of complex social outcomes into the actions and choices of the individuals who contribute to those outcomes. There is a common adherence to a *generativist* theory of social explanation—the idea that a good social explanation of a phenomenon consists in a demonstration of the circumstances through which the actions of individual actors brought it about. This approach amounts to a kind of ontological individualism—the idea that social outcomes depend on the actions and thoughts of individual actors. There is a broad, though not exceptionless, adherence to the idea that individuals are purposive and intentional in their actions. There is a strong attachment to the idea that social explanations require that we identify the causal mechanisms through which social outcomes occur, a theory that has developed rapidly within the philosophy of science in recent years as well. Finally analytical sociology is especially receptive to the methods of computational social science and agent-based modeling. By providing a model of a social arrangement (the spread of rumors, the acceleration of social protest, the proliferation of a disease through a population) based on a computational model of the interactions and actions of individual actors, the computational social scientist provides a demonstration of the mechanism through which a given outcome is generated.

1. ANALYTIC PHILOSOPHY

Let us begin by asking a key question: what are the intellectual relationships that exist between analytical sociology and analytic philosophy?

Analytic philosophy is a heterogeneous collection of research styles and approaches that are defined, primarily, by a common commitment to clarity and precision. Philosophical problems within this tradition are addressed through logical analysis and conceptual analysis. There is a strong thread of empiricism within the tradition of analytic philosophy, extending from the writings of empiricist philosophers such as Locke, Hume, and Berkeley. (Valuable surveys of the history and development of analytic philosophy are provided by Schwartz 2012 and

Daniel Little - 9781789906851

Beaney 2017.) Stephen Schwartz (2012) provides a concise description of analytic philosophy in *A Brief History of Analytic Philosophy*:

> An analytic philosopher analyzes problems, concepts, issues, and arguments. She breaks them down into their parts, dissects them, to find their important features. Insight comes from seeing how things are put together and how they can be prized apart; how they are constructed and how they can be reconstructed. Symbolic logic was and remains the most distinctive tool of analytic philosophers. (p. 2)

Analytic philosophy emerged in elite British universities in the decade preceding the First World War. It was substantively influenced by recent advances in symbolic logic and mathematics, including the work of Peano and Frege. It was also motivated by an aversion to the lack of clarity and precision that was thought to be characteristic of speculative philosophy, including, especially, the neo-Hegelians (e.g. J.M.E. McTaggart and F.H. Bradley (1897)) who were dominant within pre-war British philosophy. Logicians like Frege had undertaken to create a "language of thought" in which propositions and inferences could be expressed in perfectly exact symbolic form—what Frege called a *begriffsschrift*, or concept writing (Frege 1964). Bertrand Russell was attracted to the idea of an absolutely clear and precise form of logical expression that would permit the formulation and derivation of mathematical and conceptual truths. With Alfred North Whitehead he wrote *Principia Mathematica* (Whitehead and Russell 1910) with the goal of systematizing the advances in symbolic logic that Frege had initiated in order to complete the program of logicism (the derivation of all of mathematics from logical axioms without appeal to mathematical intuitions). The program failed for reasons that turned out to be fundamental, including the incompleteness of mathematics that Gödel proved in 1931. But this achievement itself could be regarded as a great triumph for the emerging analytic movement: a deep and important problem had been formulated, and a solution had been decisively established. There was no speculation, no conceptual vagueness and no continued debate about the original issue.

What is "analytical" about analytic philosophy? As much as any other criterion, analytic philosophy was defined by a set of intellectual attitudes and tastes. Since its emergence in the early decades of the twentieth century analytic philosophers have given great priority to linguistic and logical analysis, clarity of exposition, deductive logic, and extreme care about the meaning of philosophical and ordinary statements. Its practitioners were vigilant in the importance of adherence to standards of clarity and rigor, and they were impatient with grand, "meaningless" theories. Most fundamentally it is the *style* of reasoning that unifies philosophers in this tradition, their attempt to formulate questions clearly and then to "analyze" (dissect) their meanings and implications. Analysis has to do with discovering clearly and rigorously the constituents of a thing, whether an organism or a concept in mathematics. And often enough, analytic philosophers came to the conclusion that their analysis of the assumptions of ordinary speech and life revealed confusion and contradiction. Analytic philosophy attempted to assist thought by discovering error and building on sound foundations. There was a tendency to prefer "meta" questions rather than substantive questions (in ethics in particular). And there was a pronounced skepticism about the methods and topics of the traditional methods and topics of Continental philosophy (Hegelian idealism, existentialism, phenomenology).

Several topic areas were especially prominent in the first decades of the development of the field. These included an interest in the foundations of logic and mathematics; philosophical analysis of language; the idea of ordinary language metaphysics, or the presuppositions

Daniel Little - 9781789906851

of ordinary ways of speaking about the world; the philosophy of science, including logical positivism, verificationism, intertheoretical reduction, and the primacy of physics (Carnap 1966); meta-ethics; and the philosophy of mind and psychology (often in a behaviorist mode). Substantive metaphysics was rejected as a speculative discipline; in its place came analysis of the forms of language through which we describe the world. Substantive ethical reasoning too was suspect, and metaethics became the avenue through which the emerging movement of analytic philosophy was able to address issues of ethics and morality, with a virtually exclusive interest in the question of the meaning of ethical and value concepts.

A central topic for early analytic philosophers was the nature of language and meaning. How should we understand the fact that words have meanings, and that sentences convey thoughts? Frege (1980 [1910]) opened modern discussions of this topic with his distinction between sense and reference. The reference of a noun is the object (or class of objects) to which it refers. The sense (meaning) of the noun is the set of ideas or procedures through which the user is able to identify the referent of the word. G.E. Moore devoted a great deal of attention to the question of what is involved in understanding the meaning of a word or sentence. Wittgenstein's *Tractatus Logico-Philosophicus* (1955 [1921]) represented the fullest expression of a logicist theory of meaning, interpreting sentences as depictions of states of affairs, just as his posthumous *Philosophical Investigations* (1968) eventually came to represent a paradigmatic expression of an ordinary-language theory of meaning (meaning as language games). In the *Tractatus* Wittgenstein gave expression to an idea that eventually came to be known as the principle of significance: only propositions that can be expressed in the language of the natural sciences can be said to be meaningful. This view excludes the idea that there are meaningful statements in pure philosophy (metaphysics in particular). In *Philosophical Investigations* Wittgenstein (1968) introduced his crucial metaphors of family resemblance (as an interpretation of the meaning of some concepts), language games, and "forms of life" as a necessary context for understanding language and grammar.

This concern for the meaning of language became a core defining characteristic of analytic philosophy. Another important component also derived from a search for precision and logical rigor with regard to non-mathematical topics such as law and morality. G.E. Moore's *Principia Ethica* (1922), A.J. Ayer's *Language, Truth, and Logic* (1948), and Gilbert Ryle's *Concept of Mind* (1949) represented different ways philosophers sought to incorporate clear conceptual analysis and a commitment to theoretical minimalism into their philosophical theorizing. Ayer undertook to present the central ideas of the emerging philosophy of language associated with Russell, along with the philosophy of science associated with logical positivism, to a broader public. Ryle attempted to debunk the philosophy of mind by eliminating the necessity of referring to mental entities and constructs—the "ghost in the machine" (Ryle 1949). Ryle's book embraced a form of logical behaviorism, while also paying close attention to the particulars of ordinary language in the statements we make about our states of mind. Moore made a sustained attempt to provide clear meanings for ethical concepts and to discredit the idea that ethical concepts could be defined "naturalistically". Here is how Moore begins his treatment of the philosophy of ethics in *Principia Ethica* (1922):

> It appears to me that in Ethics, as in all other philosophical studies, the difficulties and disagreements, of which its history is full, are mainly due to a very simple cause: namely to the attempt to answer questions, without first discovering precisely what question it is which you desire to answer. (Moore 1922, xvii)

Daniel Little - 9781789906851

The burden of the book *Principia Ethica* is to provide adequate definitions of these "moral" concepts and relate these words to the topic of "good conduct". The solution to intractable philosophical problems is to be found in clear conceptual analysis of ethical language.

Another important tributary of analytic philosophy was the philosophy of science developed within the Vienna Circle. Rudolph Carnap's *Logical Structure of the World* (1967 [1928]) was offered as a minimalist conception of scientific knowledge, which eliminated speculative concepts, including theoretical hypotheses. Otto Neurath, Moritz Schlick, Friedrich Waismann and Herbert Feigl contributed to the refinement of a developed philosophy of science, logical positivism. Much of the later development of the philosophy of science developed at the Vienna School and took place in the United States, especially the work of Carnap, Hempel, Nagel, and Feigl.

Moore and Carnap are indicative of the philosophical approach taken by analytic philosophers in the founding period of the field: identify a problem area, isolate the key concepts and provide exact and organized definitions of these concepts in order to permit logical construction of answers to clear philosophical questions.

Analytic philosophy of science since the 1960s continued to evolve in ways that served to loosen the sparse minimalism of the first generation. Contemporary analytic philosophy of science is frequently characterized as "post-positivist" in the sense that many philosophers have embraced scientific realism and have rejected the dogmas of logical empiricism as a framework for understanding scientific knowledge. One important development is a post-Humean treatment of causation, opening space for a realist interpretation of causal necessity (Harré 1970). Another is the emphasis on the distinction between the context of discovery and the context of confirmation. The work of Thomas Kuhn (1970) and Imre Lakatos (1970) directed valuable attention toward the intellectual and institutional infrastructure of science, the paradigms and research programmes through which scientific research proceeds. The hypothetico-deductive method of explanation (Hempel 1965; Nagel 1961) was questioned by philosophers of social science in the 1960s and 1970s. Against the instrumentalism that characterized much positivist philosophy of science, post-positivist philosophers of science often embraced and developed the theory of scientific realism as an interpretation of the significance of scientific theories (Putnam 1982, 1984; Boyd 1984, 1990; Salmon 1984). Scientific realism has relevance in the social sciences as well as the natural and biological sciences, and the idea of real operative causal mechanisms in the social realm came to have persuasive force in the 1980s and 1990s. This field developed through fruitful engagement between philosophers, sociologists, and political scientists.

Philosophy in the English-speaking world has progressed significantly since the first expressions of analytic philosophy. The majority of professional philosophers in North America, Britain, Australia and New Zealand continue to identify as analytic philosophers, but strictures on philosophical theorizing reflecting prior commitments to formalism and positivism have loosened, and new topics are considered. Philosophy of science has been a vigorous and forward-moving discipline for at least five decades. The philosophy of psychology and neuroscience has developed into a very powerful set of theories and methods, and has largely abandoned the philosophical behaviorism that was shared by Wittgenstein and Ryle. (Notable contributions include Stich 1983, Patricia Churchland 1986, Paul Churchland 1989, Chalmers 1996, and Craver 2007.) Normative ethics and political philosophy have moved well past the "meta" approach taken by ordinary-language philosophers in the 1950s. John Rawls was immersed in the texts and methods of analytic philosophy, but his theory of

Daniel Little - 9781789906851

justice (1971) rejected the idea that ethical theories must remain at the meta-ethical level. Instead he argued for a substantive theory of justice intended to provide a basis for evaluating the justice of real social institutions. Much of the field of ethics has followed Rawls's example in offering philosophical treatments of issues of moral theory in ways that embody both analytical rigor and clarity along with substantive engagement with specific moral ideas and principles.

American pragmatism had a very distinct set of origins from British analytic philosophy. Nonetheless, the ideas of Dewey, Peirce, Mead, and James have had a profound effect on the development of American philosophy, and the intermarriage has proved fruitful. Analytic philosophy has been transformed through its encounter with American pragmatism and the writings of Charles Peirce, John Dewey, William James, and George Herbert Mead. W.V.O. Quine, Hilary Putnam, Nelson Goodman, Richard Rorty, John Rawls and Robert Brandom all show the influence of American pragmatism. Their version of analytic philosophy has incorporated the anti-foundationalism of Dewey and Peirce, and some of their most distinctive contributions derive from the union. Rawls adopts the idea of reflective equilibrium from the coherence epistemology developed by Quine, which in turn reflects the anti-foundationalism of American pragmatism. Quine's metaphor of the whole of scientific knowledge as a "web of belief" encapsulates the anti-foundationalist orientation brought to analytic philosophy by pragmatism (Quine 1960; Quine and Ullian 1970).

Does it still make sense to use the phrase "analytic philosophy" to encompass the range of practices of "anglophone philosophy" in universities in Britain, Australia, Asia and North America? Did Saul Kripke, John Rawls, Derek Parfit, Stanley Cavell, Jaegwon Kim and Tim Scanlon still have enough in common in the 1970s and 1980s to label them all "analytic" philosophers? How about Elizabeth Anderson, Peter Railton, and Derrick Darby in today's philosophy world? Certainly the fields of philosophy in English-speaking universities have gone well beyond the strictures of the narrow views of the founders, and there is a great divergence of topics and methods among these philosophers. However, this is often the case when we consider the genealogy of an intellectual field. In the end it is not the label but the content and reasoning of the theories that philosophers produce that matter. And yet, it is still correct to say that all these philosophers are more similar to each other than they are to contemporary specialists in hermeneutics, phenomenology or *assemblages* theory.

1.1 Parallels and Influences from Analytic Philosophy

Let us turn now to the possible relationships that exist between analytic philosophy and analytical sociology. What do analytic philosophy and analytical sociology have in common? Did analytic philosophy play a meaningful, substantive role in the development of analytical sociology? Does it have a valuable role to play in the future development of analytical sociology? The first is a question of intellectual influence, and it is difficult to give a definite answer. However, to judge by the references in *Dissecting the Social* (Hedström 2005), Hedström and other formative contributors to analytical sociology were well acquainted with several important segments of the field of analytic philosophy from the 1960s through 1990s. *Dissecting the Social* contains at least twenty references to important contributions in analytic philosophy during those decades, especially in the fields of the philosophy of science and the philosophy of action and psychology. That said, references to analytic philosophy are generally illustrative rather than substantive.

Daniel Little - 9781789906851

The most frequent references to analytic philosophers in *Dissecting the Social* are in the philosophy of science. In the period between 1930 and 1960 the philosophy of science was largely defined by the premises of logical positivism, the unity of science, the idea of falsifiability, and the hypothetico-deductive method. None of these themes in the philosophy of science have proven to be particularly productive for the development of the special sciences—biology, psychology, economics, sociology, or anthropology. The direct influence of positivist philosophy of science on the genesis of analytical sociology is (thankfully) very slight. Likewise, ordinary language philosophy, pure conceptual analysis, and an insistence on a sharp separation between substantive topics (metaphysics, ethics) and philosophical statements about them seem to have little echo in current work in analytical sociology. Overall, then, we might say that the influence from analytic philosophy to analytical sociology was minimal.

There are nonetheless a few substantive parallels between the two fields. First, like Hedström, analytic philosophers often turn to the metaphor of "dissection" to characterize their work (as Stephen Schwartz does in the passage quoted earlier). Here is Hedström's description of the methodology associated with analytical sociology in approaching problems for sociological explanation:

> One important characteristic of the analytical approach is that it aims to gain understanding by dissecting the social phenomena to be explained. To dissect, as the term is used here, is to decompose a complex totality into its constituent entities and activities and then to bring into focus what is believed to be its most essential elements. When focusing on what is believed to be particularly important for the problem at hand, we abstract from, or move out of focus, those elements believed to be of lesser importance. In this sense, dissection and abstraction are two aspects of the same activity, and they are core components of the analytical approach. It is through dissection and abstraction that the important cogs and wheels of social processes are made visible and intelligible. (Hedström 2005, p. 2)

Similar to analytic philosophers like Moore and Carnap, Hedström believes that clarity comes from breaking a problem down into its component parts. Having achieved a degree of decomposition of a problem, it is possible to use logic, rigorous specification of fundamental ideas, and careful argument to arrive at a better understanding of the original problem. This point highlights another similarity between analytic philosophy and analytical sociology. Both disciplines place a high value on clarity. Analytic philosophers were dismissive of the speculative and indeterminate theories of much of the tradition of philosophy in the fields of metaphysics, theology, and epistemology, because they believed that these theories could not be formulated clearly and intelligibly, and it was impossible to specify the circumstances that would allow us to conclude either in favor of or against such theories. Likewise, analytical sociologists are skeptical about the large and often poorly specified theories that have often played an important role in the history of sociology for the same reasons.

A second important parallel is the interest that both fields have in formal approaches to problem solving. As noted, analytic philosophy was profoundly affected by major advances in symbolic logic in the early twentieth century and the promise that deductive logic offered in treating problems such as meaning, truth, and mathematical proof. It is intriguing to consider whether the methodology of computational models and simulations plays an analogous role in analytical sociology to the role played by symbolic logic in analytic philosophy. In both cases the practitioners believed that the formal representation of ideas and theories in the field provided a more secure basis for analysis than ordinary language. Much of the intellectual power of computational social science derives from the aspiration of "generativism": using a model to

Daniel Little - 9781789906851

show how a complex social outcome is generated by the actions of a set of independent actors. Joshua Epstein describes agent-based models in similar terms:

> To the generativist, explaining macroscopic social regularities, such as norms, spatial patterns, contagion dynamics, or institutions requires that one answer the following question: How could the autonomous local interactions of heterogeneous boundedly rational agents generate the given regularity? Accordingly, to explain macroscopic social patterns, we generate—or "grow"—them in agent models. (Epstein 2006, p. 1)

Analytic philosophers of mathematics might well have endorsed a version of the slogan that Epstein made famous on behalf of agent-based models: "If you haven't derived your conclusions using formal symbolic logic, then you have not solved the problem".

Another important parallel between analytical sociology and analytic philosophy is the fact that neither discipline is primarily defined by a particular subject matter, but rather by a style and method of approach to problems of interest. Analytic philosophers take on a very wide range of topics; likewise, analytical sociologists are prepared to offer explanations of a wide range of puzzling sociological outcomes. In each case the research efforts are defined as "analytic" because of the methodologies that they pursue, not because of the nature of the subject matter.

There are a few areas where there has been a developed and substantive engagement between analytical sociology and analytic philosophy. Most important are on the topics of causal mechanisms and rational action. Post-positivist philosophers of science returned to the question of the meaning of causation, and a substantial critique emerged of the Humean idea that causation is nothing more than constant conjunction. Philosophers in the 1970s also came to believe that Hempel's theory of covering-law explanation (1965) did not provide an adequate analysis of causation either. One of the results of this renewed focus on causation was new attention on the idea of a causal mechanism. A realist theory of causation gained ground in analytic philosophy of science according to the idea that A causes B needs to be understood along the lines of an assertion that "there are a series of real causal mechanisms leading from the occurrence of A to the occurrence of B". This view of causation has proven to be very helpful within the philosophy of science (Elster 1989; Little 1991; Mahoney 2001; McAdam, Tarrow and Tilly 2001; Little 2011; Craver and Darden 2013). Hedström and other researchers within analytical sociology have taken the idea of causal mechanisms as a pivotal solution to the problems of explanation, attribution of causation, and heuristics of research in the social realm (Hedström and Swedberg, 1998; Hedström 2005; Hedström and Ylikoski 2010).

Another important area of discussion and theory development within analytic philosophy relevant to analytic sociology is the theory of action. Philosophers developed important critiques of excessive reliance on the idea of economic rationality, and philosophers attempted to develop theories of action that were adequate to the problem of understanding the occurrence of collective action. Donald Davidson (1963) provided an important bridge between philosophy and the social sciences with his arguments aiming to show that reasons can serve as causes of actions. Significantly, theorists from the social sciences made important contributions to these ideas within the literature of analytic philosophy, including Amartya Sen, Russell Hardin, and Jon Elster. Analytic philosophers today consider the challenges involved in the idea of group agency, collective action, and the role of the emotions in action (List and Pettit 2011).

Daniel Little - 9781789906851

1.2 Dividends from AP to AS

In the 1980s and 1990s philosophers of science underwent something of a "scientific revolution" of their own and turned to issues that were more directly relevant to the practice of scientific researchers in these various fields. Philosophers of science came to recognize the value of study of particular areas of science. There was less interest in the nature of scientific explanation, confirmation, or research in general, and more interest in identifying and contributing to problems that arise in the conduct of real areas of scientific research—biology, high-energy physics, psychology, the brain, economics, or sociology, for example. By taking on some of these immediate issues that have relevance to contemporary research, philosophers in the analytic tradition made their work more relevant and helpful for working scientists (Rosenberg 1976; Little 1993; Kincaid 1996). This is plainly true in the field of analytical sociology, where topics having to do with the nature of causal explanation, the nature of purposive action, and the nature of mid-level scientific explanation all found relevance in analytical sociology. The subfields within analytic philosophy of philosophy of biology, psychology, economics, and social science all took much greater interest in the specifics of theory and research in various fields of science, and philosophers took on the responsibility of mastering the details of the theoretical and experimental challenges that arose in those fields. Commonly there has been very productive interaction between philosophers in these fields and researchers in the corresponding areas of scientific inquiry.

There is today a productive level of interaction between philosophers of social science and practitioners of analytic sociology. Petri Ylikoski, Jon Elster, Daniel Steel, Julie Zahle and others have engaged substantively with the issues that researchers in analytical sociology are engaging. Examination of the contents of the leading journal in the philosophy of social science (*Philosophy of the Social Sciences*) confirms the extent of productive interaction that currently exists between philosophers and issues within analytical sociology. In the past three years at least half of the articles published in this journal are on topics that are of central concern to analytical sociology. Likewise the programs of the European Network for Philosophy of Social Science contain a great many papers that address the intersection between the philosophy of social science and analytical sociology.

There are several areas of fruitful intersection between the philosophy of social science and analytical sociology. Most evident are the interactions that have developed in the area of discussions of the use of causal mechanisms as a superior theory of causal explanation compared to the logical empiricist idea of the covering-law model. There is clearly a convergence of interest between sociologists and philosophers on the topic of causal mechanisms.

A second area where meaningful interactions have occurred between philosophers and analytical sociologists is in the area of the theory of action. Analytical sociology gives a great deal of attention to the logic of social action, and this requires an adequate theory of the actor. Research in analytic philosophy in the philosophy of mind and action is very relevant to this aspect of the program of analytical sociology (whether or not the promised connections bear fruit). The nature of action and intention has been an important focus of discussion in analytic philosophy (Goldman 1970; Kenny 1973; Harman 1976; Searle 1983; Dretske 1988) and a great deal of productive discussion in rationality and the assumptions of rational choice theory has occurred. Analytic philosophers have considered the role of the emotions in action and have sought to incorporate insights from pragmatist theories of action into a theory of the

Daniel Little - 9781789906851

actor. This includes philosophers who have engaged with pragmatist theories of the actor, the study of motivation and the emotions, and heuristic theories of decision-making. Also relevant is research currently underway within philosophy on the topic of group action, including Margaret Gilbert 1989; Raimo Tuomela 1995; and Christian List and Philip Pettit 2011.

A third productive area for collaboration between sociologists and philosophers is within the domain of social ontology. What is the nature of social entities? Are there "higher" and "lower" levels of social organization? What does the idea of "generativism" amount to? Is it equivalent to reductionism? Is the idea of supervenience a useful idea for social ontology or for sociologists? What does "emergence" mean? Are social properties emergent? These are topics to which philosophers of social science have devoted considerable attention, and their findings are directly relevant to topics of explanation and research within analytic sociology. These topics are all relevant to how we interpret the icon of analytical sociology, Coleman's boat (Coleman 1990).

Fourth, analytic philosophers of science have devoted attention to the issues that arise within the field of computational social science. How should we interpret a model? Does a simulation allow us to predict the future behavior of a given social system? Is a model a "realistic" description of the system it represents? How should we formulate the standards of evidence and empirical evaluation that are relevant for assessing and using computational models? This set of questions falls in the domain of what might be called the "epistemology of computational systems", and it is a matter of great importance for analytical sociology. Philosophers of science are able to make a valuable contribution to these questions, along with experts in statistics, epidemiology and other areas where computational models are especially important for research.

These points of connection demonstrate that there are very fruitful opportunities for collaboration between analytic philosophers in several subfields and researchers within the field of analytical sociology.

2. ANALYTICAL MARXISM

Let us turn now to another area of analytic thinking, analytical Marxism. There was an explosion of interest in Karl Marx by analytic philosophers in the early 1970s. Philosophers such as Allen Wood, George Brenkert, Allan Buchanan, John McMurtry, Gerald Cohen, Jon Elster, Daniel Little, Adam Przeworski (a political scientist) and John Roemer (an economist) began taking Marx's writings seriously, offering extensive analysis and criticism of his theories. Examples of some of this work are included in *Analytical Marxism: Studies in Marxism and Social Theory* (Roemer 1986) and Wright, Levine, and Sober's 1992 work. This resurgence began in discussions of "Marx's theory of justice" (e.g. Husami 1978, Cohen et al., 1980), but extended quickly into many other areas of Marx's thought—the theory of exploitation, the labor theory of value, the theory of historical materialism, and his theory of capitalism as a distinctive mode of production. Little (1986) argued for a "microfoundational" interpretation of Marx's theory of capitalism and for the idea that Marx's key explanations took the form of "institutional logic" explanations. Early arguments discrediting the labor theory of value fall in this category as well. This work was referred to as "rational choice Marxism" or "analytical Marxism", and it represented an intellectual agenda that took Marx seriously as a thinker, but readily rejected philosophical, methodological, or theoretical

Daniel Little - 9781789906851

ideas that Marx endorsed, which no longer seemed adequate to the task. This work often came to conclusions that were unacceptable to more orthodox Marxist theorists.

In anglophone philosophy prior to 1970 Marxism was generally treated roughly, criticized for its Hegelian roots and its supposed tendencies toward grand and superficial assertions. In the 1970s a new current of thinking came into academic philosophy, at least in part occasioned by the questioning of the contemporary social order occasioned by the Vietnam War, the Civil Rights movement, the anti-nuclear movement, and other protest movements against the status quo. Young philosophers who had been trained in the rigor and precision of analytic philosophy came to consider whether there were insights in Marx's writings that would be fertile in the current historical moment. Unlike many of their teachers they did not proceed from an instinctive distrust of Marxism, while at the same time having no patience with the dogmatic approaches to social theory adopted by the Communist movement since the 1930s. These philosophers undertook to bring a more appreciative approach to the reading of Marx, while maintaining their commitments to rigor, clarity, theoretical analysis and intellectual honesty. The new philosophy journal *Philosophy and Public Affairs* hosted some of the first expressions of this new receptiveness to Marxian ideas, with a particular emphasis on Marx's ideas about the justice or injustice of the capitalist system.

This new receptiveness to Marx's ideas by philosophers and colleagues in political theory quickly acquired a new name, analytical Marxism. Gerald Cohen (1978) was one of the first major contributions. Jon Elster's *Making Sense of Marx* (1985), as well as his emerging dialogue with Cohen, accelerated the movement. Adam Przeworski (1985a, 1985b), Andrew Levine (1984), John Roemer (1981), John McMurtry (1978), Erik Olin Wright (1978), Robert Brenner (1976) and Daniel Little (1986) explored multiple aspects of Marx's social theories. Piero Sraffa's *Production of Commodities by Means of Commodities* (Sraffa 1963) provided a foundation for new critical thinking about the labor theory of value. Of particular interest were topics concerning Marx's theory of historical materialism, his economic theories (including the labor theory of value), his theory of class conflict, and his ideas about alternatives to capitalism. In each case the researcher undertook to arrive at a clear and explicit interpretation of Marx's thesis; adjusted and corrected the theory; and sought to evaluate its strength in offering a basis for analysis and explanation of real phenomena in history and society. (Fabien Tarrit offers a critical analysis of the origins and development of analytical Marxism [Tarrit 2006].)

The field was sometimes referred to as "rational-choice Marxism" because of the themes that many of these investigators found in Marx, leading back to a methodology based on rational or purposive individual actors (capitalists and workers). In *The Scientific Marx* Little argued that Marx's explanations generally rested on a requirement of microfoundations for satisfactory social explanations (1986); it was argued there that Marx's central explanations of the main tendencies of capitalism (falling rate of profit, recurring crises, industrial reserve army, technological change) were based on an analysis of the interests of capitalists and workers and the strategies that each adopted. Somewhat surprisingly it was concluded that Marx's theories and explanations in *Capital* fall squarely within what we would now call "analytical sociology" with a commitment to microfoundations, mechanisms, and socially constituted purposive actors.

Analytical Marxism was not primarily a philosophical movement, but it owed a great deal to the intellectual values of clarity, precision, and analysis that analytic philosophy had championed, and there was a great deal of productive interaction between analytic philosophers (philosophy of social science; social and political philosophy; philosophy of economics) and

Daniel Little - 9781789906851

the field of analytical Marxism. Crucially, analytical Marxism provided a powerful rebuttal to the idea that there was no contemporary substance to Marx's social theories. These investigators made sense of many of Marx's central propositions about history and society. Along the way they felt entirely free to jettison the less credible parts of Marx's theories—the Hegelian language Marx sometimes adopted, Marx's adherence to the importance of the labor theory of value and Marx's ideas about inevitability in history. Most importantly these researchers helped to forge a conception of social explanation that depended on purposive actors, social structures possessing microfoundations, and a healthy respect for the contingency and path dependency of history. A particularly important contributor was Jon Elster, who was a key early contributor to analytical Marxism and has also played an important role in the development of analytical sociology. His career-long focus on problems of rationality and aggregation helped him to make important contributions to both fields, as well as to analytic philosophy.

It is very interesting in hindsight to notice how short-lived the phenomenon of analytical Marxism was. Most of the central contributors moved on from their interest in Marx's thought to other substantive issues that had little or nothing to do with Marxism. However, Diego Gambetta notes that there is a strong legacy of the core research objectives of analytical Marxism within the subsequent development of analytical sociology (Manzo 2010, p. 138). Key ideas within analytical Marxism, that have a great deal of relevance to the development of analytical sociology, include the idea that Marx's key explanations and theories are compatible with the requirement that microfoundations should be provided for macro-level claims about society; the idea that Marx's central social-scientific methods are actor-centered rather than structuralist; and the idea that the social sciences need to arrive at hypotheses about the mechanisms that underlie important social processes of interest. None of these ideas finds a comfortable home in orthodox theories of Marxism, but they were crucial to the interpretations of Marx's scientific ideas offered by researchers mentioned here. In this way analytical Marxism was indeed an important precursor and source of influence to analytical sociology.

3. ASSESSMENT

How are these analytic disciplines relevant to the emergence and contours of analytical sociology? The most important strands of intellectual kinship that exist between analytic philosophy and analytical sociology begin with the matter of intellectual style: a firm insistence on the importance of clarity and logical reasoning and a deep suspicion of grand theories—either philosophical or sociological. Both analytic philosophers and analytical sociologists seek to *dissect* the problems that they study, to uncover the components of the systems or concepts under examination.

Several specific areas of research within analytic philosophy since 1970 appear to have been especially relevant to the development of analytical sociology. Particularly important for analytical sociology is the philosophy of science. Philosophers and sociologists alike have taken a fresh interest in causal realism and causal mechanisms, and there has been active exchange between philosophers and sociologists on this topic (Hedström and Ylikoski 2010; Machamer et al., 2000; Craver and Darden 2013). Another important strand of analytic philosophy since the 1970s is the theory of action (including rational choice theory and collective action theory), and this too has great relevance for analytical sociology (Davidson 1963; Searle 1969; Elster 1979; Hardin 1982; Bratman 1987; Nozick 1993;

Daniel Little - 9781789906851

Bratman 2014). Third, the current prominence of the philosophy of pragmatism within analytic philosophy has important links to analytical sociology in that it provides intellectual resources for arriving at more fully developed theories of the actor (Rorty 1979; Joas 1993; Brandom 2008; Gross 2008; Bernstein 2010).

It is interesting to observe that Jon Elster has played an important role in all three analytic disciplines—philosophy, Marxism, and sociology. Elster's academic discipline is political science, but he has been an active contributor to analytic philosophy throughout his career and has often occupied joint appointments in both philosophy and political science. His contributions to the philosophy of social science have been of the greatest importance in philosophy. His persistent and deep interests in the complexities of rationality and imperfect rationality have been of great importance to philosophers; these areas of work have also been very important within the development of analytical sociology. Elster succeeded in offering the example to philosophers that their topics and issues should be considered in tandem with significant knowledge of the scientific disciplines that adjoin these problems. His insistence on the need for microfoundations for social explanations was grounded in analytic philosophy of science, and it played a strong role in the development of analytical sociology. His critique of functionalist explanation in his debates with G.A. Cohen (1978) demonstrated astute philosophical reasoning within the tradition of analytic philosophy, while contributing to the methodological sophistication of analytical sociology. Elster was also intimately involved in the second "analytic" field under consideration here, the field of analytical Marxism. Elster's main contributions to analytical sociology include his emphasis on the importance of causal mechanisms in social explanations and his efforts at providing a more nuanced understanding of the rational actor than is provided by traditional rational choice theory. One might say without excessive flourish that Elster (along with his French counterpart, Raymond Boudon) was one of the key connectors who brought together the theoretical and methodological insights of analytic philosophy, analytical Marxism, and analytical sociology. (Manzo 2012 provides extended discussion of Boudon's theory of action, and more detailed analysis of Elster's contributions to the theory of action can be found in Little 2014.)

These points make it clear that there are important thematic similarities between analytic philosophy and analytical sociology. However, these are substantially independent traditions. The shared devotion to clarity, logical analysis, and arguments that proceed through "dissection" do not constitute the beginnings of a hybrid field. Rather, it seems to be a common expression of a "style" of intellectual thought that emerged in the twentieth century that influenced both philosophers and social scientists. And there is very little evidence of intellectual influence in the 1960s and 1970s that would indicate that the founders of analytical sociology were modeling their ideas on those of analytic philosophy. (Exceptions include Davidson 1963; Searle 1969, 1983; and Elster 1979.) That said both analytic philosophy (in its current topics and concerns) and analytical sociology (in its future developments) can contribute much to each other.

The connections between analytic philosophy and analytical Marxism were more direct. A substantial number of the researchers who contributed to the development of the main themes of analytical Marxism received their training in graduate programs in analytic philosophy, and they brought those sensibilities to their study of Marx's theories. Moreover, there was a substantial degree of interaction between segments of the fields of analytic philosophy, American political science, and economics in American universities during those decades, and there was a high degree of shared assumptions about methodology and theory that permitted

Daniel Little - 9781789906851

easy cross-communication. When John Roemer (economics), Stephen Marglin (economics), Adam Przeworski (political science), or Andrew Levine (political science) interacted with philosophers engaged in research on topics in analytic Marxism they found a common vocabulary and very similar assumptions about epistemology that permitted easy communication. The field developed quickly, in part because of this common intellectual heritage.

The connections between analytical Marxism and analytical sociology also demonstrate an important degree of overlap. Both fields were interested in discovering the microfoundations of important social processes. Both were drawn to versions of liberalized versions of rational choice theory (desire-belief-opportunity theory, class actor theory) as a basis for social analysis and explanation. Both identified the importance of the discovery of social mechanisms as a key element in social explanations. Both focused on the idea that social explanations proceed by demonstrating how specific institutional arrangements lead actors to behaviors that aggregate to important social outcomes (institutional logic explanations, disaggregated structural explanations).

It was noted that analytic philosophy is an academic field defined primarily by its methods and style, not by a discrete family of research objectives. Both analytical Marxism and analytical sociology differ from philosophy in this regard. Neither of these "analytics" is exclusively defined by its methods; instead, both fields of research are guided by the overriding goal of providing a better basis for understanding the social world. Both are scientific research fields (Bourdieu 1996) with a core set of commitments and hypotheses to be extended and evaluated through scientific research and an extended community of researchers involved in these efforts. As scientific communities, each is well advised to observe a crucial principle of open-minded consideration of new hypotheses and approaches. Analytical Marxism was at its best when it considered new methodologies, new bodies of research, and new disciplinary perspectives that might assist its search for a fruitful framework for understanding the world of capitalist society. John Roemer's (1982) use of the methods of neoclassical economics to provide a new theory of Marxist exploitation is illustrative. This openness to alternative theories and methods was valuable for the field; but it was also crucial for the ability of analytical Marxism to draw young researchers in philosophy, economics, political science, or sociology into the field. Likewise, analytical sociology will be most successful if it can avoid the trap of dogmatism that sometimes leads to intellectual stagnation. It is crucial for the analytical sociology field to maintain its ability to communicate widely with researchers in other disciplines and to remain open to new theoretical ideas that may enhance its overall ability to explain the social world.

REFERENCES

Agassi, Joseph, and I.C. Jarvie (2019), "The problem of analytic philosophy", *Philosophy of the Social Sciences*, **49** (5), 413–33.

Ayer, A.J. (1948), *Language, Truth, and Logic*, London: V. Gollancz.

Beaney, Michael (2017), *Analytic Philosophy: A Very Short Introduction*, Oxford: Oxford University Press.

Bernstein, Richard (2010), *The Pragmatic Turn*, Cambridge: Cambridge University Press.

Bourdieu, Pierre (1996), *The Rules of Art: Genesis and Structure of the Literary Field*, Stanford, CA: Stanford University Press.

Boyd, Richard N. (1984), "The current status of scientific realism", in Jarrett Leplin (ed.), *Scientific Realism*, Berkeley, CA: University of California Press.

Boyd, Richard (1990), "Realism, approximate truth, and philosophical method", in C. Wade Savage (ed.), *Scientific Theories*, Minneapolis, MI: University of Minnesota Press.

Daniel Little - 9781789906851

Bradley, F.H. (1897), *Appearance and Reality: A Metaphysical Essay*, 2nd ed., Oxford: Clarendon Press.
Brandom, Robert (2008), *Between Saying and Doing: Towards an Analytic Pragmatism*, Cambridge, MA: Harvard University Press.
Bratman, Michael (1987), *Intention, Plans, and Practical Reason*, Cambridge, MA: Harvard University Press.
Bratman, Michael (2014), *Shared Agency: A Planning Theory of Acting Together*, New York, NY: Oxford University Press.
Brenner, Robert (1976), "Agrarian class structure and economic development in pre-industrial Europe", *Past and Present*, **70**, 30–75.
Carnap, Rudolf (1966), *Philosophical Foundations of Physics; An Introduction to the Philosophy of Science*, New York: Basic Books.
Carnap, Rudolf (1967 [1928]), *The Logical Structure of the World; and Pseudoproblems in Philosophy*, Berkeley, CA: University of California Press.
Chalmers, David John (1996), *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press.
Churchland, Patricia Smith (1986), *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, Cambridge: MIT Press.
Churchland, Paul M. (1989), *A Neurocomputational Perspective*, Cambridge, MA: MIT Press.
Cohen, G.A. (1978), *Karl Marx's Theory of History: A Defence*, Princeton: Princeton University Press.
Cohen, Marshall, Thomas Nagel and Thomas Scanlon (eds) (1980), *Marx, Justice, and History*: *A Philosophy and Public Affairs Reader*, Princeton, NJ: Princeton University Press.
Coleman, James S. (1990), *Foundations of Social Theory*, Cambridge: Harvard University Press.
Craver, Carl F. (2007), *Explaining the Brain: What the Science of the Mind-Brain Could Be*, Oxford: Oxford University Press.
Craver, Carl F. and Lindley Darden (2013), *In Search of Mechanisms: Discoveries across the Life Sciences*, Chicago: The University of Chicago Press.
Davidson, Donald (1963), "Actions, reasons, and causes", *Journal of Philosophy*, **60** (23), 685–700.
Dretske, Fred (1988), *Explaining Behavior*, Cambridge, MA: MIT Press.
Elster, Jon (1979), *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge: Cambridge University Press.
Elster, Jon (1985), *Making Sense of Marx*, Cambridge: Cambridge University Press.
Elster, Jon (1989), *Nuts and Bolts for the Social Sciences*, Cambridge: Cambridge University Press.
Epstein, Joshua (2006), *Generative Social Science: Studies in Agent-Based Computational Modeling*, Princeton, NJ: Princeton University Press.
Frege, Gottlob (1964), *Begriffsschrift und andere Aufsätze*. Hildesheim: G. Olm.
Frege, Gottlob (1980), "On sense and reference", in Peter Geach and Max Black (eds.), translated by Max Black (1980), *Translations from the Philosophical Writings of Gottlob Frege*, 3rd ed., Oxford: Basil Blackwell.
Gellner, Ernest (1979), *Words and Things: An Examination of, and an Attack on, Linguistic Philosophy* (revised), London: Routledge & Kegan Paul.
Gilbert, Margaret (1989), *On Social Facts*, Princeton NJ: Princeton University Press.
Gross, Neil (2008), *Richard Rorty: The Making of an American Philosopher*, Chicago, IL: University of Chicago Press.
Goldman, Alvin (1970), *A Theory of Human Action*, Englewood Cliffs, NJ: Prentice-Hall.
Hardin, Russell (1982), *Collective Action*, Baltimore, MD: Johns Hopkins University Press.
Harman, Gilbert (1976), "Practical reasoning", *Review of Metaphysics*, **79**, 431–63.
Harré, Rom (1970), *Principles of Scientific Thinking*, Chicago, IL: University of Chicago.
Hedström, Peter (2005), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge, UK and New York: Cambridge University Press.
Hedström, Peter and Richard Swedberg (eds) (1998), *Social Mechanisms: An Analytical Approach to Social Theory*, (Studies in Rationality and Social Change), Cambridge and New York: Cambridge University Press.
Hedström, Peter and Petri Ylikoski (2010), "Causal mechanisms in the social sciences", *Annual Review of Sociology*, **36**, 49–67.
Hempel, Carl (1965), *Aspects of Scientific Explanation, and Other Essays in the Philosophy of Science*, New York: Free Press.
Husami, Ziyad I. (1978), "Marx on distributive justice", *Philosophy and Public Affairs*, **8** (1), 27–64.
Joas, Hans (1993), *Pragmatism and Social Theory*, Chicago, IL: University of Chicago Press.
Kenny, Anthony (1973), *Action, Emotion, and Will*, London: Routledge & Kegan Paul.
Kincaid, Harold (1996), *Philosophical Foundations of the Social Sciences: Analyzing Controversies in Social Research*, Cambridge and New York: Cambridge University Press.
Kuhn, Thomas S. (1970), *The Structure of Scientific Revolutions*, 2nd ed., Chicago: University of Chicago Press.
Lakatos, Imre (1970), "The methodology of scientific research programmes", in Imre Lakatos and Alan Musgrave (eds), *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press.

Daniel Little - 9781789906851

Levine, Andrew (1984), *Arguing for Socialism: Theoretical Considerations* (*Critical and Social Thought*), Boston, MA: Routledge & Kegan Paul.
List, Christian and Philip Pettit (2011), *Group Agency: The Possibility, Design, and Status of Corporate Agents*, Oxford and New York: Oxford University Press.
Little, Daniel (1986), *The Scientific Marx*, Minneapolis, MN: University of Minnesota Press.
Little, Daniel (1991), *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science*, Boulder, CO: Westview Press.
Little, Daniel (1993), "Evidence and objectivity in the social sciences", *Social Research*, **60** (2), 363–96.
Little, Daniel (2011), "Causal mechanisms in the social realm", in Phyllis Illari, Federica Russo and Jon Williamson (eds), *Causality in the Sciences*, Oxford: Oxford University Press.
Little, Daniel (2014), "Jon Elster and the social sciences", in Claudio Lopez-Guerra and Julia Maskivker (eds), *Democracy, Rationality and Justice*, Cambridge and New York: Cambridge University Press.
Machamer, Peter, Lindley Darden, and Carl F. Craver (2000), "Thinking about mechanisms", *Philosophy of Science* **67** (1), 1–25.
Mahoney, James (2001), "Beyond correlational analysis: Recent innovations in theory and method", *Sociological Forum*, **16** (3), 575–93.
Manzo, Gianluca (2010), "Analytical sociology and its critics", *European Journal of Sociology*, **51**, 129–70.
Manzo, Gianluca (2012), "Reason-based explanations and analytical sociology: A rejoinder to Boudon", *European Journal of Sociology*, **50** (2), 35–65.
McAdam, Doug, Sidney G. Tarrow and Charles Tilly (2001), *Dynamics of Contention. Cambridge Studies in Contentious Politics*, New York: Cambridge University Press.
McMurtry, John (1978), *The Structure of Marx's World-View*, Princeton, NJ: Princeton University Press.
Moore, G.E. (1922), *Principia Ethica*, Cambridge: Cambridge University Press, 1922.
Nagel, Ernest (1961), *The Structure of Science; Problems in the Logic of Scientific Explanation*, New York: Harcourt Brace.
Nozick, Robert (1993), *The Nature of Rationality*, Princeton, NJ: Princeton University Press.
Przeworski, Adam (1985a), "Marxism and rational choice", *Politics & Society*, **14** (4), 379–409.
Przeworski, Adam (1985b), *Capitalism and Social Democracy*, Cambridge: Cambridge University Press.
Putnam, Hilary (1982), "Three kinds of scientific realism", *The Philosophical Quarterly*, **32** (128).
Putnam, Hilary (1984), "What is realism?", in Jarrett Leplin (ed.), *Scientific Realism*, Berkeley: University of California Press.
Quine, W.V. (1960), *Word and Object*. (*Studies in Communication*), Cambridge, MA: MIT Press.
Quine, W.V. and J.S. Ullian (1970), *The Web of Belief*, New York: Random House.
Roemer, John (1981), *Analytical Foundations of Marxism*, New York: Cambridge University Press.
Roemer, John (1982), *A General Theory of Exploitation and Class*, Cambridge, MA: Harvard University Press.
Roemer, John (ed.) (1986), *Analytical Marxism*, Cambridge: Cambridge University Press.
Rorty, Richard (1979), *Philosophy and the Mirror of Nature*, Princeton, NJ: Princeton University Press.
Rosenberg, Alexander (1976), *Microeconomic Law: A Philosophical Analysis*, Pittsburgh, PA: University of Pittsburgh Press.
Ryle, Gilbert (1949), *The Concept of Mind*, London: Hutchinson.
Salmon, Wesley C. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press, 1984.
Schwartz, Stephen (2012), *A Brief History of Analytic Philosophy: From Russell to Rawls*, Malden, MA and Oxford, UK: John Wiley & Sons.
Searle, John R. (1969), *Speech Acts: An Essay in the Philosophy of Language*, London: Cambridge University Press.
Searle, John R. (1983), *Intentionality: An Essay in the Theory of Mind*, New York: Cambridge University Press.
Sraffa, Piero (1963), *Production of Commodities by Means of Commodities: Prelude to a Critique of Economic Theory*, Cambridge, UK: Cambridge University Press.
Stich, Stephen P. (1983), *From Folk Psychology to Cognitive Science: The Case against Belief*, Cambridge, MA: MIT Press.
Tarrit, Fabien (2006), "A brief history, scope, and peculiarities of 'Analytical Marxism'", *Review of Radical Political Economics*, **38** (4).
Tuomela, Raim (1995), *The Importance of Us: A Philosophical Study of Basic Social Notions*, Stanford: Stanford University Press.
Whitehead, Alfred North and Bertrand Russell (1910), *Principia Mathematica*, Cambridge: Cambridge University Press.
Wittgenstein, Ludwig (1955), *Tractatus Logico-Philosophicus*, London: Routledge & Kegan Paul.
Wittgenstein, Ludwig (1968), *Philosophical Investigations*, Oxford: Blackwell.
Wright, Erik Olin (1978), *Class, Crisis and the State*, London: Verso, 1978.
Wright, Erik Olin, Andrew Levine and Elliott Sober (1992), *Reconstructing Marxism: Essays on Explanation and the Theory of History*, London and New York: Verso.

Daniel Little - 9781789906851

3. Analytical sociology and its lesser-known antecedents: structural individualism in European sociology[1]

Werner Raub

1. INTRODUCTION

Antecedents of analytical sociology are mentioned and discussed in programmatic pieces and manifestos, such as the introductory essay (Hedström and Swedberg 1998b) for *Social Mechanisms: An Analytical Approach to Social Theory* (Hedström and Swedberg 1998a); Hedström's (2005) *Dissecting the Social. On the Principles of Analytical Sociology*; the chapters by Hedström and Bearman (2009b) and Hedström and Udehn (2009) in the 'Foundations' part of *The Oxford Handbook of Analytical Sociology* (Hedström and Bearman 2009a), and Manzo's (2010, 2014b) detailed overall assessments of analytical sociology.

Antecedents include work by classical sociologists, specifically Weber and Tocqueville (see for example, Hedström 2005, p. 6 for further references), as well as more contemporary work. Merton's (1968) discussion of middle-range theories and his examples of such theories are clearly considered prime antecedents (Hedström and Udehn 2009; see also Hedström and Swedberg 1998b; Hedström 2005; Hedström and Bearman 2009b). An important antecedent is Coleman's (1990) micro-macro approach to explaining macro-level phenomena and regularities, as summarized in his well-known diagram (see for example, Hedström and Swedberg 1998b, pp. 21–3; Hedström 2005: Chapter 6 and passim; Hedström and Udehn 2009, pp. 32–3; see Figure 3.1).[2] At times (see for example, Hedström 2005, Chapter 3; Hedström and Ylikoski 2010, p. 60), it is stressed, too, that Coleman's micro-macro approach can and should be distinguished from a related feature of his research agenda, namely, employing rational choice theory as a micro theory of behavior, furthermore arguing that rational choice theory should better *not* be employed as an 'ingredient' of analytical sociology.

Manzo (2010, pp. 133–6) has shown how analytical sociology is related to earlier work on 'mechanisms' in theory construction and statistical modeling. With respect to theory construction, Manzo emphasizes early work in mathematical sociology (for example, Coleman 1964) that can be seen as aiming at specifying formal models of mechanisms. Somewhat related, Hedström elaborates on the study of mechanisms in his discussions of sociological explanations (see Hedström 2005: Chapter 2; Hedström and Bearman 2009b, p. 5), typically referring to earlier work by Elster (1989). Related to the link between mechanisms and statistical modeling, both Hedström (for example 2005, pp. 110–13) and Manzo emphasize that the study of mechanisms is closely connected to providing formal theory in the sense of 'generative

[1] Gianluca Manzo's very useful comments on an earlier version are gratefully acknowledged.

[2] Note the dedication of Hedström's and Swedberg's (1998a: p. v) edited volume: 'For James S. Coleman and Robert K. Merton', highlighting the importance of Coleman and Merton for analytical sociology. The 1998 volume did not use the expression 'analytical sociology' as such, only 'an analytical approach to social theory'. Still, it might be seen, together with the 1996 Stockholm conference that led to the volume, as the 'go-ahead' for analytical sociology as an explicit 'approach' (Hedström 2005, p. 1) to theory and empirical research.

Werner Raub - 9781789906851

models'. Employing clearly specified assumptions, such models imply statistical regularities. Empirical data can then be used to estimate relevant parameters of the models. In this respect, the study of mechanisms builds on earlier work, such as Coleman (1981) and Cox (1992).

With respect to theoretical analyses of social phenomena, Schelling's (1971) model of racial segregation is perhaps the foremost example of an antecedent for analytical sociology (for example, Hedström 2005, pp. 7–8; Hedström and Bearman 2009b, pp. 12–13). Schelling's analysis is typically also mentioned as an early example of agent-based models, with such models as a key tool for theory construction in analytical sociology (for example, Hedström 2005; Manzo 2010, 2014a). Granovetter's (1978) threshold models of collective behavior are another paradigm often mentioned as having been influential for the development of analytical sociology (for example, Hedström and Bearman 2009b, p. 13).[3] Both Schelling's and Granovetter's models carefully link micro- and macro-levels of analysis, exemplifying a core feature of what analytical sociology tries to achieve. Their models also focus on how actors are interrelated and interdependent and on the consequences of interrelatedness and interdependence for social outcomes, with social outcomes often being unintended by the actors involved and often also hard to predict for the researcher, certainly so when analysis of the micro level is neglected. This too is a core theme of analytical sociology (for example, Hedström and Bearman 2009b, p. 13).

Concerning empirical work, research on social networks should be added to our inventory (Hedström 2005, passim; Hedström and Bearman 2009b; Manzo 2014b). Given the overall aim of analytical sociology to bridge the gap between quantitative empirical research and sociological theory (Hedström 2005, p. 8), it is not surprising that antecedents in the broad field of social network analysis specifically include work linking empirical findings and regularities systematically with theoretical assumptions that can contribute to providing explanations, such as Granovetter's (1973, 1974) work on weak ties or the Coleman, Katz and Menzel (1966) study on the diffusion of innovation.

The antecedents mentioned so far are well known. Moreover, the available literature elaborates in some detail how analytical sociology builds on those antecedents. Interestingly, still another antecedent is mentioned rather regularly (for example, Hedström 2005, p. 5 n4; Hedström and Bearman 2009b, p. 8), namely, 'structural individualism'. The label refers to work in the 1970s and 1980s by a group of sociologists mainly affiliated with universities in Germany and the Netherlands, as well as work by Boudon in France. This antecedent, though, is typically not discussed in any detail, at times referring exclusively to a section in Udehn's (2001, pp. 318–9) study of different strands of methodological individualism, without references to primary sources (Manzo 2007 could well be the only analytical sociology paper with some additional information on and discussion of structural individualism). Boudon's is a special case because he contributed to the Hedström and Swedberg (1998a) volume (Boudon 1998). Boudon is mentioned for at least two contributions to the development of analytical sociology. One is his micro-macro approach to sociological explanations *combined with* his skepticism concerning rational choice theory (Hedström 2005, p. 7; see, for example, Boudon 1998). The other is his *Education, Opportunity, and Social Inequality* (Boudon 1974) that can

[3] See Hegselmann (2017) for a detailed account of how the attention for Schelling's model is somewhat biased, as well as how Schelling's and Granovetter's models are related.

Werner Raub - 9781789906851

be seen as an empirically focused study already employing some of the heuristic guidelines that inform analytical sociology (Manzo 2014b, p. 9 n2). More information on other contributions of structural individualism is largely missing in the analytical sociology literature and how structural individualism relates to analytical sociology thus remains opaque.

In this chapter, we show that key features of analytical sociology are indeed rather closely related to structural individualism. Based on original literature from the 1970s and 1980s, the chapter provides a sketch of structural individualism, with a focus on selected topics and key features that are antecedents of analytical sociology. Our sketch shows that 'structural individualism' in the sense of the present chapter, that is, as it developed in the 1970s and 1980s, indeed includes a variant of methodological individualism, as adequately brought to the fore by Udehn (2001). However, the present chapter likewise shows that 'structural individualism' in fact refers to a considerably broader set of notions than exclusively a variant of methodological individualism.

2. SOME BACKGROUND

Founders and key contributors of structural individualism in Germany include Karl-Dieter Opp, Hans J. Hummell, Rolf Ziegler, and somewhat later Hartmut Esser (see Opp 2009, pp. 27–30 for a brief sketch, including a personal account).[4] Hans Albert played an important role as a philosopher, specifically through his contributions to critical rationalism in the spirit of Popper, and through his work on the intersection of philosophy and the social sciences. Viktor Vanberg showed how structural individualism builds on the history of ideas in the social sciences and was also conducive in making some of Coleman's work better known in sociology in Germany. Wolfgang Sodeur and Franz Urban Pappi were more loosely associated with structural individualism. In the Netherlands, Siegwart Lindenberg and Reinhard Wippler have been major driving forces, over time joined more and more by Frans Stokman. Raymond Boudon's work in France in the 1970s was another key source of structural individualism.

Some of the founders and key contributors in Germany were trained at the University of Cologne, where René König and later Erwin K. Scheuch—both not connected themselves to structural individualism—had established, in the 1950s and 1960s, the 'Cologne School' of empirical social research, including methods of social research, and had secured important infrastructure for such research. Later, the University of Mannheim became another center where structural individualism flourished. In the Netherlands, Utrecht University and the University of Groningen hosted sociology departments that developed into centers of structural individualism. At the University of Groningen in the 1960s, Ivan Gadourek, like König and Scheuch, not himself involved in developing structural individualism, had become one of the founders of modern empirical social research and research methods in the Netherlands. Wippler, trained at the University of Groningen with Gadourek as his thesis advisor, became the intellectual leader of the department at Utrecht University and, likewise, an able institution builder soon after his Utrecht appointment in 1971. Lindenberg, trained at the University of

[4] It goes without saying that the following list is selective and some omissions may be unfortunate. Students of the founders and key contributors are not mentioned here, although quite a few of these students contributed themselves in important ways to structural individualism.

Werner Raub - 9781789906851

Mannheim and at Harvard, became a strong intellectual influence in Groningen after his arrival in the early 1970s. In the early 1970s, they launched an active 'inter-university workgroup in theoretical sociology'. Somewhat later, in the 1980s Wippler and Lindenberg, together with colleagues, likewise founded the Interuniversity Center for Social Science Theory and Methodology (ICS), a collaboration of Utrecht University and the University of Groningen. Nijmegen University joined the ICS later as did, much more recently, a group of sociologists at the University of Amsterdam. The ICS swiftly developed into a highly visible and influential research center and a model for PhD training in social science, not only for the Netherlands but also internationally. From the beginning, structural individualism strongly influenced the ICS research and training agenda.

Much of the original literature on structural individualism from the 1970s was published in German, Dutch and French—at the time, publishing in English was less common for scholars in those countries. Still, it is useful to note that the founders and key contributors of structural individualism, as well as their students, always have been active and visible internationally and soon built up a strong network of international contacts. For example, Ziegler and Hummell, later joined by Sodeur, organized a series of yearly international conferences on mathematical sociology (MASO) at the Werner–Reimers-Foundation in Bad Homburg near Frankfurt a.M. in 1974–1982 (see Mayer n.d. for some details), including a 'grey series' of proceedings (often, revised versions of MASO-papers appeared later in very visible outlets). These conferences were meeting points of well-established and younger sociologists from Germany and the Netherlands, with leading colleagues from the US and elsewhere, including, for example, Peter M. Blau, Raymond Boudon, James Coleman, Mark Granovetter, Michael Hannan, Gudmund Hernes, Karl G. Jöreskog, Edward Laumann, Samuel Leinhardt, Peter Marsden, Anatol Rapoport, Aage Sørensen, Harrison White, and others. It is tempting to consider these meetings a 'prototype' for the analytical sociology conferences that started in 2008 and evolved into the INAS conferences.

From a history of ideas perspective, structural individualism is a research program related to distinct traditions in the social sciences and philosophy of science (see, for example, Diekmann and Voss 2018 for a concise overview). Vanberg (1975) and Bohnen (1975) provided detailed studies on the roots of structural individualism. These roots are in the work of the Scottish Moralists of the 18th century as represented by David Hume and Adam Smith. Other sources are the Austrian School of Economics, including Carl Menger, Joseph Schumpeter, Mises, and later Friedrich Hayek, the work of Max Weber, and various strands of methodological individualism, including Karl Popper, JohnWatkins and Joseph Agassi (see also the reader O'Neill 1973).

From sociology, work by George C. Homans (1950, 1958, 1961 and 1964) and Peter M. Blau (1964) was influential for the development of structural individualism. Homans's 'behavioral sociology' had applied theories from (social) psychology in microsociology, and he had initiated the development of social exchange theory. Blau was another founder of social exchange theory in modern sociology. Still another important source for structural individualism was James Coleman's (1964) work on mathematical sociology. Merton's (1968) approach to social theory should presumably be added.

Finally, structural individualism built on contributions from the philosophy of science. This is reflected in the emphasis on methodological principles of social science explanation and testing social science theories in line with Popper's 'critical rationalism' as well as analytical philosophy, broadly conceived.

Werner Raub - 9781789906851

3. STRUCTURAL INDIVIDUALISM: KEY FEATURES AND TOPICS ANTICIPATING ANALYTICAL SOCIOLOGY[5]

Key features of structural individualism (the label might have been used first in Opp 1978a, pp. 33–4 and in the programmatic sketch Wippler 1978a, with Wippler 1985 as an 'update' some years later[6]; see Opp 1979 and Raub and Voss 1981 for more detailed early overviews, and Raub 1982 for an early edited volume) include underlying methodological principles of social science explanation and theory formation as well as the role of methodological individualism and micro-macro links. Due to methodological individualism, the specification of micro-level assumptions on individual behavior becomes an important issue. Key features furthermore include heuristic guidelines for micro-macro modeling, such as a focus on how interdependencies between actors affect individual behavior on the micro level as well as the macro-level consequences of behavior, together with a focus on how interdependence often yields unintended macro-level consequences. Model building and at least a certain degree of formalization of theories and explanations are key features, including early variants of agent-based models. Structural individualism typically aims at what is commonly referred to as middle-range theories, including careful specification of assumptions and implications of such theories. Finally, linking theoretical and empirical work is typical for structural individualism, with important studies particularly on social networks. This section briefly reviews these key features and topics, also indicating relations to later work in analytical sociology, together with some remarks on how structural individualism evolved after the 1970s and 1980s.[7]

3.1 Some Philosophy of Science Principles Underlying Structural Individualism

More than other strands in sociology, structural individualism is explicitly linked to work in philosophy of science. Structural individualism conceives of sociology as a problem- and theory-guided discipline. Theory construction aims at the explanation of social phenomena, including well-established empirical regularities. Explanations involve deductive arguments or variants of such arguments. Therefore, theory construction involves more than specifying sets of hypotheses. Rather, theory construction comprises specifying assumptions, including but not exclusively hypotheses, as well as specifying implications of these assumptions. Due to the focus on implications, analytical rigor is an ingredient of structural individualism. Implications should include *testable* implications: empirical content in the sense of testability (at least in principle) is a criterion for appraising sociological theories. Likewise, empirical tests of implications are a core aim and empirical corroboration is a criterion for appraising theories in addition to testability. Thus, the integration of theory construction and empirical

---

[5] Some parts of this section overlap with parts of Raub and Voss (2017) on structural individualism, but the present overview is often somewhat more detailed. The remarks on the relations between structural individualism and analytical sociology are of course new.

[6] The title of Wippler (1985) refers to 'explanatory sociology', another label sometimes used for structural individualism.

[7] Since much of the original literature is not in English, it seems useful to reference explicitly some sources published in English in the 1970s and 1980s or more recently that provide further information and against which at least parts of the current chapter can be checked. Early sources include Lindenberg (1985, 1992), Opp (1988), the chapters in Raub (1982), Wippler (1978a, 1985), and Wippler and Lindenberg (1987). More recent sources include Raub and Voss (2017), as well as some sections in Kroneberg and Kalter (2012) and Raub (2017).

Werner Raub - 9781789906851

research is an aim too. The affinity with a meta-theoretical approach such as Popper's (1934, 1963) and Lakatos's (1970) is evident. Indeed, Albert, who had been influential in advocating Popper's philosophy of science in Germany (Albert 1968) and in advocating Popperian principles also for social science theory formation and empirical research, contributed in important ways to the methodological foundations of structural individualism (e.g. Albert 1967, 1977). Also, analyses of the structure of scientific theories and explanations as provided by Hempel (1965) and Nagel (1961) have been clearly influential for structural individualism, including work by Carnap and others (for example, Carnap and Stegmüller 1959) on 'explication'. Opp (1970, with numerous subsequent editions) and Esser et al. (1977) are textbooks on philosophy of science in this vein for sociologists.

Structural individualism resonates well with analytical sociology in several respects related to philosophy of science. This includes, but is by far not restricted to, an explicit interest in and discussion of philosophy of science issues as those indicated, although this alone already distinguishes both structural individualism and analytical sociology from many other intellectual enterprises in sociology. What is more, the focus of analytical sociology on 'precision and clarity' (for example, Hedström 2005, pp. 1, 3–4; see also the other programmatic contributions to analytical sociology mentioned) hardly differs from structural individualism, but differs quite a bit, as Hedström (2005, p. 4) has forcefully argued, from much other 'theorizing' in sociology. The same holds for systematically linking theory and empirical research, hence the importance of testability and empirical corroboration.

Philosophy of science has evolved since the 1960s and 1970s. For example, Hempel's model of scientific explanation seems to be less of a standard model by now. Indeed, this is reflected in a rather fierce critique by Ylikoski (2013) of a defense of Hempel's model in Opp's (2013a) discussion of analytical sociology (see also Opp's 2013b reply to Ylikoski). This chapter is on antecedents of analytical sociology. Therefore, the chapter does not aim at contributing to an assessment of the relative merits of various models of scientific explanation. Also, the chapter does not contribute to the ongoing discussion on how to adequately conceive of '(causal) mechanisms' and 'mechanism-based explanations' (for example, Hedström and Ylikoski 2010) that are often presented as a core feature of analytical sociology. Irrespective of how to adequately explicate the concept of 'scientific explanation' and, more specifically, how to explicate special features of explanations in sociology, both structural individualism and analytical sociology share conceptualizing explanations as answers to why questions (for example, Hedström 2005, p. 2) and, therefore, as being different from descriptions as well as from concept formation (for example, Hedström 2005, Chapter 2). Moreover, a shared perspective is that explanations can be reconstructed as deductive arguments (see Hedström and Udehn 2009, p. 44n19 for an explicit formulation), that doing so is fruitful and that such arguments are an important feature of sound theory formation. Again, structural individualism and analytical sociology are not only similar to each other in these respects, they also differ very much from what is common in many other approaches in the discipline (for further discussion on explanation, see the next sub-section on methodological individualism and micro-macro links).

3.2 Methodological Individualism and Micro-Macro Links

In line with methodological individualism, structural individualism aims at explanations that include two levels; namely, the macro-level of social phenomena and processes, including

Werner Raub - 9781789906851

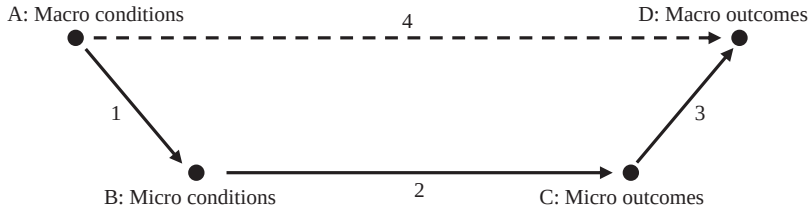


Figure 3.1 Coleman's diagram

well-established statistical regularities, and the micro-level of individuals and individual behavior. Moreover, and more particularly, structural individualism focuses on how both levels are related (Hummell 1973; Lindenberg 1977, 1981; Opp 1979; Raub and Voss 1981). Meanwhile, Coleman's (1986, 1990) micro-macro diagram is commonly used for summarizing the 'logic' of such explanations (Figure 3.1): Macro-outcomes (Node D in Coleman's diagram) and macro-regularities (Arrow 4) represent sociological explananda. Macro-conditions (Node A) affect the micro-conditions for individual behavior (Node B). Arrow 1 in Coleman's diagram represents assumptions that specify how macro-conditions affect micro-conditions for individual behavior. Node C represents micro-outcomes and the explananda at the micro-level such as individual behavior, with Arrow 2 representing micro-theories, such as more or less general theories of individual behavior. Finally, Arrow 3 represents assumptions on how actors' behavior generates macro-outcomes, thus linking the micro- with the macro-level.

Coleman's micro-macro explanations and his diagram have been rather influential for analytical sociology. In fact (see references mentioned in the introduction), Coleman's diagram is often presented as reflecting how explanations are conceived in analytical sociology. The diagram is even used to elaborate a key concept in discussions of explanations as envisaged by analytical sociology, namely social mechanisms and to distinguish different types of mechanisms. In their programmatic introduction, Hedström and Swedberg (1998b, pp. 21–3) distinguish between three types of mechanisms and relate each one to one of the arrows in Coleman's diagram. First, situational mechanisms, according to Hedström and Swedberg (1998b, p. 23), 'link a social structure or other macro-sociological event to the beliefs, desires, and opportunities of some individual actor'. Such mechanisms are specified in the assumptions represented by Arrow 1 in Coleman's diagram. Second, action-formation mechanisms are 'located at the micro-level' and show 'how a specific combination of individual desires, beliefs, and action opportunities generate a specific action' (Hedström and Swedberg 1998b, p. 23). Arrow 2 in Coleman's diagram represents assumptions on such mechanisms. Hedström and Swedberg (1998b) also mention examples: 'General decision theories as well as more specific theories such as […] cognitive dissonance and […] on discounting illustrate different types of action mechanisms.' Finally, Hedström and Swedberg distinguish transformational mechanisms. These cover 'the micro-to-macro transition […] Here a number of individuals interact with one another, and the specific mechanism (which differs depending on the nature of the interaction) shows how these individual actions are transformed into some kind of collective outcome' (Hedström and Swedberg 1998b, p. 23). Of course, Hedström and Swedberg relate transformational mechanisms to Coleman's Arrow 3 that represents assumptions on micro-to-macro transitions.

Werner Raub - 9781789906851

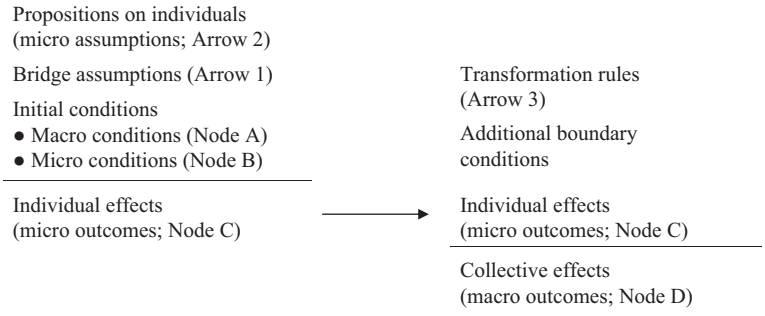


Figure 3.2 Lindenberg's (1976, 1977) model (including references to the related nodes and arrows in Coleman's diagram)

It is quite striking that the 'logic' of sociological explanations, in line with Coleman's diagram, had been specified years earlier in structural individualism, including various diagrams that are predecessors of Coleman's (for details and extensive references, see Raub and Voss 2017; see Manzo 2007, pp. 12–15 for an analytical sociology paper that does refer to some of these predecessors and Barbera 2006, p. 44n15 for a remark in the same direction). Presumably the most clear-cut case is Lindenberg's analysis of structural individualistic explanations.[8] In a series of articles in the 1970s, Lindenberg developed a model with exactly the components of Coleman's diagram, albeit organized in a somewhat different way, namely adapting the Hempel-Oppenheim model (Hempel 1965) of explanations (the major source is Lindenberg 1977; see also Lindenberg 1976, 1981, and Lindenberg and Wippler 1978). Figure 3.2 depicts Lindenberg's model. To facilitate comparison, we slightly expand his diagram by adding where the various elements (nodes and arrows) from Coleman's diagram are located in Lindenberg's version. Note that the horizontal lines indicate that the proposition below (the explanandum) is implied by the assumptions above (the explanans). Figure 3.2 shows that micro-macro explanations involve two steps. In the first, micro-outcomes (individual effects) are explained using a micro-theory together with initial conditions comprising assumptions on macro- as well as micro-conditions and bridge assumptions. In the second step, macro-outcomes ('collective effects') are derived from a set of assumptions that include the micro outcomes from the first step, possibly additional boundary conditions and the transformation rules.

Summarizing, explananda at the micro level follow from an explanans comprising assumptions on individual behavior (Node B, Arrow 2), macro-conditions (Node A), and bridge assumptions (Arrow 1). The explananda at the macro-level, that is, macro-outcomes (Node D) and macro-regularities (Arrow 4), follow from an explanans comprising assumptions on individual behavior (Node B, Arrow 2), macro-conditions (Node A), as well as bridge assumptions (Arrow 1) and transformation rules (Arrow 3). Therefore, in structural individualism sociological explanations focus on macro-phenomena as explananda and try to highlight macro-conditions rather than exclusively micro-conditions as part of the explanation.

[8] See Raub and Voss (2017, pp. 19–28) for other antecedents of Coleman's diagram in structural individualism, such as Hummell and Opp (1968) and Boudon (1979b).

Werner Raub - 9781789906851

Subsequent work elaborated Lindenberg's analysis and contributed to clarifying characteristics of micro-macro explanations (see Opp 1979; Raub 1984, Part I; Wippler and Lindenberg 1987; somewhat later Lindenberg 1992; Esser 1993; see Raub, Buskens and Van Assen 2011; Opp 2009, 2011 for more recent discussions and further references). Note, too, that Lindenberg's analysis shows that micro-macro explanations in sociology can be seen as special cases of explanations in the spirit of the Hempel-Oppenheim model. Also, the discussion of three types of social mechanisms in Hedström and Swedberg (1998b) is along rather similar lines, like Lindenberg's (1977, 1981) discussion of bridge assumptions and transformation rules, as well as Esser's (1993, pp. 93–8 and more broadly Chapters 5 and 6) discussion of the assumptions represented by the arrows in Coleman's diagram. It stands to reason that the examples discussed by Lindenberg and Esser are cases of social mechanisms in the sense of Hedström and Swedberg. Partially similar terminology already indicates this, like 'transformation rules' (Lindenberg) and 'transformational mechanisms' (Hedström and Swedberg). Esser suggested the labels 'logic of the situation' for Coleman's Arrow 1, 'logic of selection' for Arrow 2, and 'logic of aggregation' for Arrow 3. Note the similarity with 'situational mechanisms', later used by Hedström and Swedberg, with Popper's 'situational logic' (1945) as an obvious common source. More generally, explanations following Lindenberg's scheme in Figure 3.2 show how macro-outcomes and macro-regularities result from social mechanisms in the sense of Hedström and Swedberg (1998b; see Kalter and Kroneberg 2014 for further discussion).

3.3 Micro-Level Theories

We have seen that structural individualism aims at providing explanations of macro-level phenomena and processes that include assumptions on macro-micro- as well as micro-macro-links as well as assumptions on micro-level behavior. This induced a strong interest in general theories of behavior that yield assumptions represented by Arrow 2 in Coleman's diagram. It also induced work of a more methodological nature on characteristics of theories of behavior that make them *useful* for sociological explanations as envisaged by structural individualism. After all, such explanations aim at explananda not only on the micro-level but also, and particularly, on the macro-level. Hence, it becomes important that assumptions on micro-level behavior can be combined with bridge assumptions as well as transformation rules so that macro-level implications do indeed follow. Wippler and Lindenberg (1987) and Lindenberg (1985) are examples of methodological analyses that clarify these issues.

With respect to theories of behavior, structural individualism, influenced by behavioral sociology à la Homans, initially focused on theories from (social) psychology. In a careful and detailed study Hummell (1969) discussed learning theories as developed in behaviorism, as well as cognitive theories, including cognitive learning theories and theories of cognitive consistency, such as, for example, the theory of cognitive dissonance. Hummell likewise outlined sociological applications of such psychological theories. Opp (1972) is a monograph on behavioral sociology in a similar spirit. Many applications of these psychological theories in sociology focused on small group phenomena. Over time and seemingly induced at least in part by trying to apply structural individualism also in other research domains of sociology, the focus shifted to cognitive theories of motivational psychology, like variants of expectancy-value theories, and again somewhat later to variants of rational choice theories (see, for example, Opp 1978a and 1978b for early attempts in this direction). Many contributions to structural individualism continued to employ rational choice assumptions, including game

Werner Raub - 9781789906851

theory (see Diekmann and Voss 2018 for an overview). Note, however, that 'theoretical pluralism' has been characteristic for structural individualism when it comes to micro-assumptions on regularities of behavior: it would be a misunderstanding to consider rational choice theory, let alone a specific variant of that theory, as the only micro-level theory consistent with, or the only theory actually employed in, structural individualism. (Manzo 2014b, pp. 21–7 makes basically the same point with respect to micro-assumptions in analytical sociology.) Both Lindenberg (for example, 2001, 2013), with his theory of social rationality, and Esser (for example, Esser and Kroneberg 2015; Kroneberg 2014), with his model of frame selection, later developed micro-level theories that deviate in various respects from standard rational choice assumptions. It should be evident that theories of behavior employed in structural individualism are cases of action-formation mechanisms in the sense of Hedström and Swedberg (1998b). Moreover, some of these theories could be seen as providing specifications of *how* desires, beliefs and opportunities in the sense of Hedström's (2005) DBO theory jointly drive individual actions (see Diekmann 2010; Opp 2013a; Kalter and Kroneberg 2014). And, finally, theories of behavior used in structural individualism are related to assumptions on behavioral regularities used in agent-based models.

3.4 Heuristics for Micro-Macro Modeling: Interdependencies, Unintended Consequences of Individual Behavior, Formal Model Building

Contributors to structural individualism developed various heuristic guidelines for micro-macro modeling. At least three of these are closely related to proposals that later became prominent in analytical sociology. First, related to useful assumptions for micro-macro explanations, structural individualism employed the idea to search for interdependencies between actors that affect individual behavior and drive macro outcomes. This includes, but need not be restricted to, interdependence in the sense that the outcome of an actor's behavior depends not only on the actor's own choices and possible chance events, but also on the behavior of other actors and *vice versa*.

Second, due to interdependence, macro outcomes are often unintended consequences of individual behavior: the very fact that outcomes also depend on the behavior of others means that outcomes of behavior, including macro outcomes, need not correspond to actors' intentions, even if behavior does depend on intentions in one way or the other. For example, Hummell (1973) carefully analyzed how interdependence and micro- as well as macro-effects of interdependence can be modeled. Boudon (1977) and Wippler (1978b, 1981) provided important work on unintended consequences from the perspective of structural individualism. Once again the focus of analytical sociology on social interaction as well as on 'structural interdependence' (Manzo 2014b, pp. 27–9) as a guiding principle for theory construction is clearly in line with structural individualism.

Third, formal model building or at least formalizing certain assumptions—and providing arguments why doing so is useful—is a further feature of structural individualism. The MASO conferences have been mentioned already and indicate a key topic. It is closely related to issues arising from aiming at micro-macro explanations as well as from the focus on interdependencies as a core ingredient of such explanations. Full-fledged micro-macro explanations typically comprise a sizeable number of possibly complex assumptions, certainly so when such explanations account for interdependencies. Deriving implications from these assumptions or deciding whether a given explanandum does or does

Werner Raub - 9781789906851

not follow from the assumptions in the explanans is often a non-trivial task. In structural individualism authors like Ziegler (1972) and Hummell (1972, 1973) elaborated the rationale for formal micro-macro models. They stressed that the core issue of model building is not the translation as such of informal verbal theory into some formal language. Rather, in terms of micro-macro explanations, the core issue related to formalization is different. Namely, the core issue is to employ well-specified formal language *and associated rules for deriving consequences from given premises* in order to be able to check for the implications of assumptions on macro and micro conditions, as well as bridge assumptions and transformation rules for micro and, more particularly, macro-outcomes, including macro regularities. An early version of this approach can be found in Coleman's *Introduction to Mathematical Sociology* and more specifically in his discussion of 'synthetic theories' (Coleman 1964, Chapters 1.4 and 18). The major aim of synthetic theories is to identify the consequences of a set of assumptions and more precisely to identify consequences on the macro-level for a set of assumptions on the micro-level: 'it is characteristic of many of these theories that they begin with postulates on the individual level and end with deductions on the group level' (Coleman 1964a, p. 41).

Ziegler and Hummell showed in detail that formal theoretical models allow identification assumptions that are actually used in deriving certain conclusions and that all too often remain implicit in purely informal accounts. Based on case studies with careful and systematic formal reconstructions of micro-macro models, they showed, furthermore, that formal model building is often necessary to be at all able to derive macro-consequences from micro assumptions. The title of Hummell (1973) refers to 'methodological individualism, structural effects, and consequences for social systems'. Coleman's diagram is already encapsulated in this title, with methodological individualism quite obviously related to Arrow 2, structural effects related to bridge assumptions (Arrow 1), and consequences for social systems related to transformation rules (Arrow 3). The title conveys that formal model building is useful for making sure that micro-macro models allow for deriving implications on the macro level, often including at first sight counterintuitive implications. Hummell (1973, p. 66) remarks that, from the perspective of methodological individualism, sociology should not be equated with (social) psychology, but rather with 'social psychology plus mathematics'. The remark is provocative and—as Hummell acknowledges explicitly—invites misunderstandings. Nevertheless, Hummell accentuates the core idea, provided one conceives of social psychology broadly enough so that it includes a broad range of theories of individual behavior on the micro-level; not exclusively theories commonly associated with psychology as a discipline, and so that it likewise includes assumptions on social conditions and interdependencies between actors.

Formal model building typically faces a trade-off. On the one hand simplifying assumptions preserves tractability and analytical power, allowing for the derivation of implications, including testable implications, at the cost of being less realistic. On the other hand more complex and realistic assumptions make it more difficult to derive implications. Addressing this trade-off, Lindenberg (1992) suggested to employ the 'method of decreasing abstraction'. The idea is to start with a model that is as simple as possible, making simplifying assumptions explicit. Subsequently, one can introduce more complex assumptions in a stepwise fashion when simplifying assumptions turns out to be problematic because, for example, implications are highly dependent on such assumptions rather than being robust or because implications fare badly in light of empirical evidence.

Werner Raub - 9781789906851

It seems clear that the structural individualistic perspective on formal model building is closely related to the 'quest for precision and clarity' (Hedström 2005, p. 3) in analytical sociology. Consider one of Manzo's (2014a, p. 7, P3) principles of analytical sociology: 'in order to explain the social outcome(s) described, first formulate a "generative model"; that is, a model of a (set of) mechanism(s) where a mechanism is a set of entities and activities likely to trigger a sequence of events (i.e. a process) likely to bring about the outcome(s)'. It stands to reason that formal model building as set forth in structural individualism by Ziegler, Hummell and Lindenberg indeed yields assumptions on such mechanisms and allows for deriving implications that describe the respective social outcomes.

3.5 Antecedents of Agent-Based Modeling

Agent-based modeling is often mentioned as a key theoretical tool of analytical sociology (see Hedström 2005; Manzo 2010, 2014b). This is reflected in the actual use of agent-based modeling in applications (see, for example, various chapters in Manzo 2014a and in Demeulenaere 2011) as well as in an overview chapter (Macy and Flache 2009) in Hedström and Bearman (2009a). Agent-based modeling cannot be claimed as a core feature of structural individualism. However, there is a link between structural individualism and analytical sociology related to agent-based modeling. Boudon's (1974) *Education, Opportunity, and Social Inequality* is an influential study on social mobility and the sociology of education, applying structural individualism in research on key empirical problems of sociology. His IEO- and ISO-models of inequality of educational opportunity (IEO) and inequality of social opportunity (ISO) are no full-fledged agent-based models. Still, they can be seen as agent-based models *in statu nascendi* in the sense of analyzing (not to say simulating) the effects of only one, or at best very few, parameter constellations for macro-outcomes related to educational and social inequalities and, particularly, trends in such inequalities. It deserves mentioning that this antecedent of analytical sociology has been repeatedly highlighted (Hedström 2005, pp. 112–3; Manzo 2014b, p. 9; see Manzo 2013 for detailed discussion), also making it clear that Boudon's models, for all their shortcomings, are early sociological examples of generative theoretical models that imply observed statistical regularities (see also Boudon 1979a).

3.6 Reconstruction of Middle Range Theories

A recurrent feature in contributions to structural individualism are reconstructions (very much in the spirit of explication à la Carnap) of well-known sociological theories and specifically of so-called middle range theories. This work aims to show that reconstructing influential middle range theories by specifying their assumptions, including implicit assumptions, carefully and explicitly, often as formal theoretical models, does in fact yield micro-macro models of social phenomena and processes in line with principles of structural individualism. Often, another aim is showing that such reconstructions allow for deriving new implications and predictions, or for showing that certain implications only hold under certain additional assumptions that had been overlooked in earlier work.

To mention only some examples, reconstructions of middle range theories are core building blocks of Boudon's (1979b) textbook and to some degree of his collection of essays (Boudon 1977). His case studies include theories of relative deprivation, Sombart's analysis of why there is no socialism in the US, theories of intra-organizational processes, theories of

Werner Raub - 9781789906851

educational and social inequalities, Michels' 'iron law of oligarchy', Schelling's models of segregation and diffusion models. Many of Boudon's reconstructions are more or less informal sketches. Wippler (1986) has provided a rather detailed reconstruction of Michels's iron law of oligarchy. Formal theoretical models based on various middle range theories are introduced and discussed in Hummell (1973) and Ziegler (1972). These include various theories of group formation, interaction in small groups, diffusion, opinion formation in groups, and executive compensation in organizations.

Analytical sociology aims at middle range theories and is skeptical with respect to the chances of successfully developing more general theories in sociology (Hedström and Udehn 2009, p. 26 and passim). Structural individualism contributed to careful specification of middle range theories and, importantly, provided specifications of such theories in terms of micro-macro models. In this respect too, structural individualism is an antecedent of analytical sociology. It is true, too, that Opp, one of the founders of structural individualism in the 1970s and 1980s, more recently offered a critical discussion of middle range theories and their role in analytical sociology (Opp 2013a, pp. 347–53). Opp's key point, though, is not that it would be misguided to develop and improve middle range theories. His key point seems that work on middle range theories would likewise benefit from developing and employing more general micro-level theories of behavior. Such micro-level theories—Opp (for example, 2013a, pp. 340–7) envisages his 'wide version' of rational choice theory—could be used as a common core in various middle range theories (see Diekmann and Voss 2004, p. 20 for an earlier discussion of this idea). Thus, the shared focus of structural individualism and analytical sociology on middle range theories is consistent with divergent assessments of whether such a common core, let alone rational choice theory as a common core, should be aimed at and would be needed (on this, see the exchange Ylikoski 2013, pp. 391–2 and Opp 2013b, pp. 406–7). As has been highlighted, structural individualism does indeed aim at explanations that include micro-level assumptions on behavioral regularities. However, structural individualism does not presuppose a *specific* behavioral theory such as a version of rational choice theory. Moreover, while the use of various versions of rational choice assumptions became rather common in the later work of many contributors to structural individualism, certainly in the early phase of the development of structural individualism in the 1970s quite different behavioral assumptions were employed. In this respect too, analytical sociology resembles structural individualism rather closely.

3.7 A Focus of Empirical Research: Social Networks

Structural individualism, again similar to analytical sociology, is aimed at closely linking theory construction and empirical research. Many contributors to structural individualism indeed engaged themselves in theoretical as well as empirical work. Structural individualism has induced empirical research in a broad range of different domains of sociology, too much to be covered in this chapter (see Diekmann and Voss 2018 for some pointers to further reading). With an eye on structural individualism as an antecedent of analytical sociology, though, one specific field of empirical research should be mentioned, namely, social network analysis. Already in the 1970s much empirical work in structural individualism has been on social networks. (Ziegler 2010 offers an overview of empirical network studies in Germany, including collaboration with network researchers in the Netherlands and elsewhere in Europe; see Fennema and Heemskerk 2018 for a somewhat related survey from a Dutch perspective).

Werner Raub - 9781789906851

Major examples of empirical network studies include a series of projects on interlocking directorates, culminating in a comparative analysis of ten countries (Stokman et al. 1985). Hummell and Sodeur (1984) analyzed the dynamics of the networks of relations in student populations. Pappi has been involved in important work on community influence systems (for example, Laumann and Pappi 1976), and Pappi and Melbeck (1984) report on one such study in Germany. A series of empirical studies on the creation of and returns on social capital in very different contexts have been led or advised over a long period of time by Flap and his group (see Flap and Völker 2013 for an overview).

With respect to such empirical research, the commonality between structural individualism and analytical sociology is more than research on social networks. The commonality includes that network effects are addressed as well as the emergence and dynamics of networks. It is interesting to note that long before network effects and network formation became fashionable topics in economics (see, for example, Bramoullé et al., 2016), Hummell (1973, p. 127, my translation) observed in the concluding section of his study on formal micro-macro models: 'The most interesting approach would be to treat networks not as exogenously given conditions but to explain network formation itself in the framework of the theory to be applied. Networks and their characteristics would then become intervening variables to be explained, using the theory and initial conditions, and would themselves affect system level consequences.'

3.8 Developments Since the 1980s

This chapter mainly, though not exclusively, sketched structural individualism in the 1970s and 1980s. Since then substantive as well as institutional developments are noteworthy. Concerning substantive developments it is quite obvious that employing rational choice assumptions as a theoretical core of micro-macro models has become more dominant over time (for overviews, see Wittek et al., 2013; Raub, forthcoming). This development has led to using rather different versions of such assumptions and to lively discussions on the pros and cons of different versions. There are versions similar to those employed in neoclassical economics (for example, Braun 1993). Applications of game theory, including variants such as behavioral and experimental game theory as well as evolutionary game theory, are other examples (see Buskens and Raub 2013 for an overview). Still other examples include wide versions of rational choice theory as proposed by Opp, as well as efforts that can be seen as aiming at a general theory of action that would include more standard rational choice assumptions as special cases. Esser's model of frame selection (Esser and Kroneberg 2015; Kroneberg 2014) and Lindenberg's (2001, 2013) theory of social rationality are examples of work in this direction. Still, a common feature of otherwise, in many respects, different applications of rational choice assumptions—and one that differs from how analytical sociology is usually presented—seems to be the idea of employing rational choice assumptions, in whatever version, as a common core of different micro-macro models.

Other substantive developments concern empirical research as well as linking theory building and empirical research, including statistical modeling. These include more emphasis on applications of experimental designs in addition to observational studies (see, for example, Buskens and Raub 2013; Raub 2017). They also include developing and employing statistical models that facilitate linking theoretical models and results of empirical research by better integrating core theoretical assumptions into the statistical models themselves. (Wittek et al., 2013 offer examples as well pointers to relevant further literature.) It

Werner Raub - 9781789906851

seems clear that these developments are nicely in line with analytical sociology (see, for example, various chapters in Hedström and Bearman 2009a and Manzo 2014a).

From an institutional perspective, structural individualism has been thriving since the 1980s in various respects. For example, the ICS has been influential, not only in sociology in the Netherlands, through the research conducted, but by setting standards for training graduate students, and by successful placement of ICS PhDs in academic positions. It seems fair to say that this contributed to less fragmentation of the discipline, more agreement on common standards for research and teaching and thus more fertile ground for growth of knowledge in sociology in the Netherlands. In Germany, similar developments have taken place at the University of Mannheim, and many sociologists trained in the tradition of structural individualism have secured academic positions, including influential ones, at numerous universities in German-speaking countries. A recent development is the foundation of the *Akademie für Soziologie* in 2017. This is a new professional organization based on aims and principles that resemble those of structural individualism in various respects. Finally, it goes without saying that the publication culture has changed quite radically since the 1970s and 1980s, with ever more focus on publications in English and publications in international journals and other international outlets as the new standard.

4. CONCLUSION: STRUCTURAL INDIVIDUALISM AS ANALYTICAL SOCIOLOGY *AVANT LA LETTRE*?

Our overview has shown that it is not without reason, upon closer inspection, that every now and then the analytical sociology literature mentions structural individualism as an antecedent. We have seen, too, that it is an antecedent in various respects and related to core features of analytical sociology. Of course, there are common sources too, such as Merton (1968) and Coleman (1964).

The literature on structural individualism from, roughly, the 1970s and 1980s has been at center stage in our overview for two reasons. It includes key contributions, while much of this literature has been published in other languages than English, and hence has been less accessible for readers not versed in these languages. This language issue has another twist. The shift to publishing in English has been accompanied by a focus on readers, including readers in non-English speaking countries. Due to this, much of the literature from the 1970s and 1980s, surveyed in this chapter, is by now less visible even in those academic communities where it was produced.[9] Note, though, that this lack of visibility should be put in perspective. First, structural individualism and features like those outlined in this chapter have become quite influential and have affected various intellectual and institutional developments as well as diverse research fields in sociology. It seems quite in line with structural individualism itself to emphasize growth of knowledge and how certain ideas contributed to the growth of knowledge, rather than too much emphasis on history of ideas in the sense of presenting

[9] A lighter note is tempting in this respect. In his *Sociology as a Population Science* that likewise resembles structural individualism in some respects, Goldthorpe (2016, p. 6n4) remarks: 'A reviewer of a paper that a similarly aged colleague and I recently submitted to a leading sociological journal objected to the fact that articles were cited that were published before he or she was born. Clearly, nothing of importance could have happened before that data'. In our case, this problem is further complicated by the language issue.

Werner Raub - 9781789906851

earlier work for its own sake. Second, it is obvious that many key contributors to structural individualism became highly visible and influential, also internationally, in the discipline. This has been due to the quality of their work, outstanding international networks of scholarly contacts and international activities, and because they themselves adapted their publication strategy, routinely publishing in English, frequently in very visible outlets. Often, this work built on ideas like those sketched in this chapter, thus also contributing to the diffusion of those ideas in the discipline.

Finally, structural individualism as analytical sociology *avant la lettre*. Science and the growth of knowledge tend to benefit from pluralism combined with cumulative growth of knowledge, as opposed to fragmentation. This includes theoretical pluralism in the sense of alternative theories and explanations for the same empirical phenomena and regularities, empirical pluralism in the sense of testing the same theory in various domains and also pluralism in the sense of testing the same hypotheses using alternative and complementary research designs, such as observational studies and survey research as well as experiments.

Analytical sociology 'seeks to explain complex social processes by carefully dissecting them and bringing into focus their most important constituent components. The approach focuses on traditional sociological concerns but uses explanatory strategies more often found in analytical philosophy and behavioral economics. It is an approach that seeks precise, abstract, realistic and action-based explanations for various social phenomena' (Hedström 2005, p. 1).

Given this, one way of conceiving of analytical sociology would be as a broad research program. In this way analytical sociology could comprise various variants of micro-macro modeling. Similarly, one would then conceive of DBO theory as a useful umbrella term for alternative and competing micro-level theories. Current competitors would be variants of rational choice models as indicated, as well as variants of, for example, models assuming myopic behavior, backward-looking learning models (with respect to backward-looking learning models and also models assuming myopic behavior, Hummell 1969 could be considered remarkably up to date), and evolutionary models, including analytical models as well as computational simulation models, with analytical sociology and DBO theory as (part of) the common core and focus of these models. Thus conceived, analytical sociology could play an important role in bringing to the fore common ground between different and competing approaches to rigorous theory and empirical research in sociology. In this way, analytical sociology could mitigate rather than further contribute to fragmentation of the discipline.

Hedström and others seemingly prefer another and considerably more specific positioning of analytical sociology. This includes accentuating the differences between mechanism-based explanations and explanations in Hempel's sense (for example, Hedström 2005, Chapter 2; Hedström and Bearman 2009b; Hedström and Ylikoski 2010). It includes an interpretation of DBO theory such that rational choice theories and learning theories become alternatives to rather than special cases of DBO theory (for example, Hedström 2005, p. 41). It further includes to conceive of computational simulation models as a key building block of analytical sociology, sometimes suggesting that employing analytical models is excluded when doing analytical sociology (for example, Hedström and Bearman 2009b, p. 16; however, see Manzo 2014b, pp. 8–9, 29–34, 40 for a nuanced discussion of this issue). Conceived in this way, analytical sociology becomes a competitor for, say, micro-macro models employing variants of rational choice assumptions. Analytical sociology even becomes a competitor for *any* approach that aims at employing the same assumptions on micro-level regularities of behavior in micro-macro models that otherwise seek to explain explananda in different fields of the discipline.

Werner Raub - 9781789906851

Also, analytical sociology becomes a competitor for approaches aiming at analytical rather than simulation models.

A careful discussion of pros and cons of conceiving analytical sociology in one way or the other, and how the discipline could benefit best from analytical sociology, seems useful but is beyond the scope of this chapter. Still, a tentative conclusion could be that structural individualism would be even more analytical sociology *avant la lettre* under the broader interpretation of what analytical sociology is aiming at.

REFERENCES[10]

Albert, Hans (1967), *Marktsoziologie und Entscheidungslogik [The Sociology of Markets and the Logic of Decision]*, Neuwied: Luchterhand.

Albert, Hans (1968), *Traktat über kritische Vernunft [Tract on Critical Reason]*, Tübingen: Mohr.

Albert, Hans (1977), 'Individuelles Handeln und soziale Steuerung. Die ökonomische Tradition und ihr Erkenntnisprogramm [Individual action and social steering. The economic tradition and its research program]', in Hans Lenk (ed.), *Handlungstheorien—interdisziplinär IV: Sozialwissenschaftliche Handlungstheorien und spezielle systemwissenschaftliche Ansätze*, München: Fink, pp. 177–225.

Barbera, Filippo (2006), 'A star is born? The authors, principles and objectives of analytical sociology', *Revista de Sociologia*, **80**, 31–50.

Blau, Peter M. (1964), *Exchange and Power in Social Life*, New York, NY: Wiley.

Bohnen, Alfred (1975), *Individualismus und Gesellschaftstheorie [Individualism and Social Theory]*, Tübingen: Mohr.

Boudon, Raymond (1974), *Education, Opportunity, and Social Inequality*, New York, NY: Wiley.

Boudon, Raymond (1977), *Effets pervers et ordre social*, Paris: PUF. English translation: *The Unintended Consequences of Social Action*, London: Macmillan, 1982.

Boudon, Raymond (1979a), 'Generating models as a research strategy', in Robert K. Merton, James S. Coleman and Peter H. Rossi (eds), *Qualitative and Quantitative Social Research. Papers in Honor of Paul F. Lazarsfeld*, New York, NY: Free Press, pp. 51–64.

Boudon, Raymond (1979b), *La logique du social*, Paris: Hachette. English translation: *The Logic of Social Action*, London: Routledge 1981.

Boudon, Raymond (1998), 'Social mechanisms without black boxes', in Peter Hedström and Richard Swedberg (eds), *Social Mechanisms. An Analytical Approach to Social Theory*, Cambridge, UK: Cambridge University Press, pp. 172–203.

Bramoullé, Yann, Andrea Galeotti and Brian W. Rogers (eds) (2016), *The Oxford Handbook of the Economics of Networks*, Oxford, UK: Oxford University Press.

Braun, Norman (1993), *Socially Embedded Exchange*, Frankfurt a.M.: Lang.

Buskens, Vincent and Werner Raub (2013), 'Rational choice research on social dilemmas', in Rafael Wittek, Tom A.B. Snijders and Victor Nee (eds), *Handbook of Rational Choice Social Research*, Stanford, CA: Stanford University Press, pp. 113–50.

Carnap, Rudolf and Wolfgang Stegmüller (1959), *Induktive Logik und Wahrscheinlichkeit [Inductive Logic and Probability]*, Wien: Springer.

Coleman, James S. (1964), *Introduction to Mathematical Sociology*, New York, NY: Free Press.

Coleman, James S. (1981), *Longitudinal Data Analysis*, New York, NY: Basic Books.

Coleman, James S. (1986), 'Social theory, social research, and a theory of action', *American Journal of Sociology*, **91** (6), 1309–35.

Coleman, James S. (1990), *Foundations of Social Theory*, Cambridge, MA: Belknap Press.

Coleman, James S., Elihu Katz and Herbert Menzel (1966), *Medical Innovation: A Diffusion Study*, Indianapolis, IN: Bobbs Merrill.

Cox, David R. (1992), 'Causality: some statistical aspects', *Journal of the Royal Statistical Society, Series A*, **155 (2)**, 291–301.

[10] By its nature, this chapter has quite some references. As a service to the reader, non-English titles have been translated. Also, where versions in English are available, we refer to these versions rather than to (similar) versions in another language, even if those other versions have appeared earlier.

Werner Raub - 9781789906851

Demeulenaere, Pierre (ed.) (2011), *Analytical Sociology and Social Mechanisms*, Cambridge, UK: Cambridge University Press.
Diekmann, Andreas (2010), 'Analytische Soziologie und Rational Choice [Analytical sociology and rational choice]', in Thomas Kron and Thomas Grund (eds), *Die Analytische Soziologie in der Diskussion*, Wiesbaden: VS Verlag, pp. 193–204.
Diekmann, Andreas and Thomas Voss (2004), 'Die Theorie rationalen Handelns. Stand und Perspektiven [The theory of rational action. State of the art and perspectives]', in Andreas Diekmann and Thomas Voss (eds), *Rational Choice Theorie in den Sozialwissenschaften. Probleme und Perspektiven. Rolf Ziegler zu Ehren*, München: Oldenbourg, pp. 13–29.
Diekmann, Andreas and Thomas Voss (2018), 'Rational-Choice-Rezeption in der deutschsprachigen Soziologie [Rational choice in sociology in Germany, Austria, and Switzerland]', in Stephan Moebius and Andrea Ploder (eds), *Handbuch Geschichte der deutschsprachigen Soziologie*, Wiesbaden: Springer VS, pp. 663–82.
Elster, Jon (1989), *Nuts and Bolts for the Social Sciences*, Cambridge, UK: Cambridge University Press.
Esser, Hartmut (1993), *Soziologie: Allgemeine Grundlagen [Sociology: General Foundations]*. Frankfurt a.M.: Campus.
Esser, Hartmut, Klaus Klenovits and Helmut Zehnpfennig (1977), *Wissenschaftstheorie [Theory of Science]*, two volumes, Stuttgart: Teubner.
Esser, Hartmut and Clemens Kroneberg (2015), 'An integrative theory of action: the model of frame selection', in Edward J. Lawler, Shane R. Thye and Jeongkoo Yoon (eds), *Order on the Edge of Chaos*, Cambridge, UK: Cambridge University Press, pp. 63–85.
Fennema, Meindert and Eelke M. Heemskerk (2018), 'When theory meets methods: the naissance of computer assisted corporate interlock research', *Global Networks*, **18** (1), 81–104.
Flap, Henk and Beate Völker (2013), 'Social capital', in Rafael Wittek, Tom A.B. Snijders and Victor Nee (eds), *Handbook of Rational Choice Social Research*, Stanford, CA: Stanford University Press, pp. 220–51.
Goldthorpe, John (2016), *Sociology as a Population Science*, Cambridge, UK: Cambridge University Press.
Granovetter, Mark S. (1973), 'The strength of weak ties', *American Journal of Sociology*, **78**, 1360–80.
Granovetter, Mark S. (1974) *Getting a Job*, Cambridge, MA: Harvard University Press.
Granovetter, Mark S. (1978), 'Threshold models of collective behavior', *American Journal of Sociology*, **83**, 1420–43.
Hedström, Peter (2005), *Dissecting the Social. On the Principles of Analytical Sociology*, Cambridge, UK: Cambridge University Press.
Hedström, Peter and Peter Bearman (eds) (2009a), *The Oxford Handbook of Analytical Sociology*, Oxford, UK: Oxford University Press.
Hedström, Peter and Peter Bearman (2009b), 'What is analytical sociology all about?', in Peter Hedström and Peter Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford, UK: Oxford University Press, pp. 3–24.
Hedström, Peter and Richard Swedberg (eds) (1998a), *Social Mechanisms. An Analytical Approach to Social Theory*, Cambridge, UK: Cambridge University Press.
Hedström, Peter and Richard Swedberg (1998b), 'Social mechanisms: an introductory essay', in Peter Hedström and Richard Swedberg (eds), *Social Mechanisms. An Analytical Approach to Social Theory*, Cambridge, UK: Cambridge University Press, pp. 1–31.
Hedström, Peter and Lars Udehn (2009), 'Analytical sociology and theories of the middle range', in Peter Hedström and Peter Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford, UK: Oxford University Press, pp. 25–47.
Hedström, Peter and Petri Ylikoski (2010), 'Causal mechanisms in the social sciences', *Annual Review of Sociology*, **36**, 49–67.
Hegselmann, Rainer (2017), 'Thomas C. Schelling and James M. Sakoda: the intellectual, technical, and social history of a model', *Journal of Artificial Societies and Social Simulation*, **20** (3), 158 pp.
Hempel, Carl G. (1965), *Aspects of Scientific Explanation*, New York, NY: The Free Press.
Homans, George C. (1950), *The Human Group*, London: Routledge.
Homans, George C. (1958), 'Social behavior as exchange', *American Journal of Sociology*, **63** (6), 597–606.
Homans, George C. (1961), *Social Behavior. Its Elementary Forms*, New York, NY: Harcourt, rev. ed. 1974.
Homans, George C. (1964), 'Bringing men back in', *American Sociological Review*, **29** (5), 809–18.
Hummell, Hans J. (1969), 'Psychologische Ansätze zu einer Theorie sozialen Verhaltens [Psychological approaches to a theory of social behavior]', in René König (ed.), *Handbuch der empirischen Sozialforschung*, Vol. II, Stuttgart: Enke, pp. 157–277.
Hummell, Hans J. (1972), 'Zur Problematik der Ableitung in sozialwissenschaftlichen Aussagensystemen [On the problem of deduction in social science systems of propositions]', *Zeitschrift für Soziologie*, **1**, 31–46, 118–38.
Hummell, Hans J. (1973), 'Methodologischer Individualismus, Strukturefekte und Systemkonsequenzen [Methodological individualism, structural effects, and system level consequences]', in Karl-Dieter Opp and Hans J. Hummell (eds), *Probleme der Erklärung sozialer Prozesse II: Soziales Verhalten und soziale Systeme*, Frankfurt a.M.: Athenäum, pp. 61–134.

Werner Raub - 9781789906851

Hummell, Hans J. and Karl-Dieter Opp (1968), 'Sociology without sociology. The reduction of sociology to psychology: a program, a test, and the theoretical relevance', *Inquiry*, **11**, 205–26.
Hummell, Hans J. and Wolfgang Sodeur (1984), 'Interpersonelle Beziehungen und Netzstruktur. Bericht über ein Projekt zur Analyse der Strukturentwicklung unter Studienanfängern [Interpersonal relations and network structure. Report on a project on the analysis of the development of social structure among a population of new students]', *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, **36**, 511–56.
Kalter, Frank and Clemens Kroneberg (2014), 'Between mechanism talk and mechanism cult: new emphases in explanatory sociology and empirical research', *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, **66**, 91–115.
Kroneberg, Clemens (2014), 'Frames, scripts, and variable rationality: an integrative theory of action', in Gianluca Manzo (ed.), *Analytical Sociology. Actions and Networks*, Chichester, UK: Wiley, pp. 97–123.
Kroneberg, Clemens and Frank Kalter (2012) 'Rational choice theory and empirical research: methodological and theoretical contributions in Europe', *Annual Review of Sociology*, **38**, 73–92.
Lakatos, Imre (1970), 'Falsification and the methodology of scientific research programmes', in Imre Lakatos and Alan Musgrave (eds), *Criticism and the Growth of Knowledge*, Cambridge, UK: Cambridge University Press, pp. 91–196.
Laumann, Edward O. and Frans Urban Pappi (1976), *Networks of Collective Action: A Perspective on Community Influence Systems*, New York, NY: Academic Press.
Lindenberg, Siegwart (1976), 'De struktuur van theorieën van collectieve verschijnselen [The structure of theories of collective phenomena]', in Wil Aarts, Siegwart Lindenberg and Reinhard Wippler (eds), *Gedrag en struktuur*, Rotterdam: Universitaire Pers Rotterdam, pp. 1–20.
Lindenberg, Siegwart (1977), 'Individuelle Effekte, kollektive Phänomene und das Problem der Transformation [Individual effects, collective phenomena, and the problem of transformation]', in Klaus Eichner and Werner Habermehl (eds), *Probleme der Erklärung sozialen Verhaltens*, Meisenheim: Hain, pp. 46–84.
Lindenberg, Siegwart (1981), 'Erklärung als Modellbau [Explanation as model building]', in Werner Schulte (ed.), *Soziologie in der Gesellschaft*, Bremen: Zentraldruckerei der Universität, pp. 20–35.
Lindenberg, Siegwart (1985), 'An assessment of the new political economy: its potential for the social sciences and for sociology in particular', *Sociological Theory*, **3** (1), 99–114.
Lindenberg, Siegwart (1992), 'The method of decreasing abstraction', in James S. Coleman and Thomas J. Fararo (eds), *Rational Choice Theory. Advocacy and Critique*, Newbury Park, CA: SAGE, pp. 3–20.
Lindenberg, Siegwart (2001), 'Social rationality versus rational egoism', in Jonathan H. Turner (ed.), *Handbook of Sociological Theory*, New York, NY: Kluwer, pp. 635–68.
Lindenberg, Siegwart (2013), 'Social rationality, self-regulation, and well-being', in Rafael Wittek, Tom A.B. Snijders and Victor Nee (eds), *Handbook of Rational Choice Social Research*, Stanford, CA: Stanford University Press, pp. 72–112.
Lindenberg, Siegwart and Reinhard Wippler (1978), 'Theorienvergleich: Elemente der Rekonstruktion [Comparing theories: elements for reconstruction]', in Karl-Otto Hondrich and Joachim Matthes (eds), *Theorienvergleich in den Sozialwissenschaften*, Darmstadt: Luchterhand, pp. 219–31.
Macy, Michael W. and Andreas Flache (2009), 'Social dynamics from the bottom up. Agent-based models of social interaction', in Peter Hedström and Peter Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford, UK: Oxford University Press, pp. 245–68.
Manzo, Gianluca (2007), 'The progress and "urgency" of modeling in sociology. The concept of the "generative model" and its implementation' (English version of the original article in French), *L'Année Sociologique*, **57**, 13–61.
Manzo, Gianluca (2010), 'Analytical sociology and its critics', *Archives Européennes de Sociologie*, **51** (1), 129–70.
Manzo, Gianluca (2013), 'Educational choices and social interactions: a formal model and a computational test', *Class and Stratification Analysis. Comparative Social Research*, **30**, 47–100.
Manzo, Gianluca (ed.) (2014a), *Analytical Sociology. Actions and Networks*, Chichester, UK: Wiley.
Manzo, Gianluca (2014b), 'Data, generative models, and mechanisms: more on the principles of analytical sociology', in Gianluca Manzo (ed.), *Analytical Sociology. Actions and Networks*, Chichester, UK: Wiley, pp. 4–52.
Mayer, Karl Ulrich (n.d.), 'Die Arbeitsgruppe "Mathematik in den Sozialwissenschaften" (MASO) in der Werner-Reimers-Stiftung 1974–1982', *mimeo*, 6 pp.
Merton, Robert K. (1968), *Social Theory and Social Structure*, rev. and enl. ed., Glencoe, IL: The Free Press.
Nagel, Ernest (1961), *The Structure of Science*, London: Routledge.
O'Neill, John (ed.) (1973), *Modes of Individualism and Collectivism*, London: Heinemann.
Opp, Karl-Dieter (1970), *Methodologie der Sozialwissenschaften*, Reinbek: Rowohlt, 7th ed., Wiesbaden: Springer VS 2014.
Opp, Karl-Dieter (1972), *Verhaltenstheoretische Soziologie [Behavioral Sociology]*, Reinbek: Rowohlt.
Opp, Karl-Dieter (1978a), *Theorie sozialer Krisen [Theory of Social Crises]*, Hamburg: Hoffmann und Campe.
Opp, Karl-Dieter (1978b), 'Das "ökonomische Programm" in der Soziologie [The "economic program" in sociology]', *Soziale Welt*, **29**, 129–54.
Opp, Karl-Dieter (1979), *Individualistische Sozialwissenschaft [Individualistic Social Science]*, Stuttgart: Enke.

Werner Raub - 9781789906851

Opp, Karl-Dieter (1988), 'The individualistic research program in sociology', in Gerard Radnitzky (ed), *Centripetal Forces in the Sciences, Vol. II*, New York: Paragon House, pp. 208–24.
Opp, Karl-Dieter (2009), 'Das individualistische Erklärungsprogramm in der Soziologie. Entwicklung, Stand und Probleme [The individualistic research program in sociology. Development, present state, and problems]', *Zeitschrift für Soziologie*, **38** (1), 26–47.
Opp, Karl-Dieter (2011), 'Modeling micro-macro relationships: problems and solutions', *Journal of Mathematical Sociology*, **35** (1–3), 209–34.
Opp, Karl-Dieter (2013a), 'What is analytical sociology? Strengths and weaknesses of a new sociological research program', *Social Science Information*, **53** (3), 329–60.
Opp, Karl-Dieter (2013b), 'Rational choice theory, the logic of explanation, middle-range theories and analytical sociology: a reply to Gianluca Manzo and Petri Ylikoski', *Social Science Information*, **53** (3), 394–408.
Pappi, Franz Urban and Christian Melbeck (1984), 'Das Machtpotential von Organisationen in der Gemeindepolitik [The power potential of organizations in community politics]', *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, **36**, 557–84.
Popper, Karl (1934), *Logik der Forschung*, 5th ed., Tübingen: Mohr 1973. English translation: *The Logic of Scientific Discovery*, London: Hutchinson 1959.
Popper, Karl (1945), *The Open Society and Its Enemies*, 5th rev. ed., London: Routledge 1966.
Popper, Karl (1963), *Conjectures and Refutations*, London: Routledge.
Raub, Werner (ed.) (1982), *Theoretical Models and Empirical Analyses*, Utrecht: ESP.
Raub, Werner (1984), *Rationale Akteure, institutionelle Regelungen und Interdependenzen [Rational Actors, Institutional Rules, and Interdependencies]*, Frankfurt a.M: Lang.
Raub, Werner (2017), *Rational Models*, Utrecht: Utrecht University.
Raub, Werner (forthcoming), 'Rational choice theory in the social sciences', in Markus Knauff and Wolfgang Spohn (eds), *The Handbook of Rationality*, Cambridge, MA: MIT Press.
Raub, Werner, Vincent Buskens and Marcel A.L.M. van Assen (2011), 'Micro-macro links and microfoundations in sociology', *Journal of Mathematical Sociology*, **35** (1–3), 1–25.
Raub, Werner and Thomas Voss (1981), *Individuelles Handeln und gesellschaftliche Folgen [Individual Actions and Societal Level Implications]*, Darmstadt: Luchterhand.
Raub, Werner and Thomas Voss (2017), 'Micro-macro models in sociology: antecedents of Coleman's diagram', in Ben Jann and Wojtek Przepiorka (eds), *Social Dilemmas, Institutions, and the Evolution of Cooperation*, Berlin: De Gruyter, pp. 11–36.
Schelling, Thomas C. (1971), 'Dynamic models of segregation', *Journal of Mathematical Sociology*, **1**, 143–86.
Stokman, Frans, Rolf Ziegler and John Scott (eds) (1985), *Networks of Corporate Power. A Comparative Analysis of Ten Countries*, Cambridge, UK: Polity Press.
Udehn, Lars (2001), *Methodological Individualism*, London: Routledge.
Vanberg, Viktor (1975), *Die zwei Soziologien. Individualismus und Kollektivismus in der Sozialtheorie [The Two Sociologies. Individualism and Collectivism in Social Theory]*, Tübingen: Mohr.
Wippler, Reinhard (1978a), 'The structural-individualistic approach in Dutch sociology: toward an explanatory social science', *Netherlands Journal of Sociology*, **14** (2), 135–55.
Wippler, Reinhard (1978b), 'Nicht-intendierte soziale Folgen individueller Handlungen [Unintended social consequences of individual actions]', *Soziale Welt*, **29**, 155–79.
Wippler, Reinhard (1981), 'Erklärung unbeabsichtigter Handlungsfolgen: Ziel oder Meilenstein soziologischer Theoriebildung? [Explaining unintended consequences of individual actions: aim or milestone of theory formation in sociology?]', in Joachim Matthes (ed.), *Lebenswelt und soziale Probleme*, Frankfurt a.M.: Campus, pp. 246–61.
Wippler, Reinhard (1985), 'Explanatory sociology: the development of a theoretically oriented research programme', *Netherlands Journal of Sociology*, **21** (2), 63–74.
Wippler, Reinhard (1986), 'Oligarchic tendencies in democratic organizations', *Netherlands Journal of Sociology*, **22** (1), 1–17.
Wippler, Reinhard and Siegwart Lindenberg (1987), 'Collective phenomena and rational choice', in Jeffrey C. Alexander, Bernhard Giesen, Richard Münch and Neil J. Smelser (eds), *The Micro-Macro Link*, Berkeley, CA: University of California Press, pp. 135–52.
Wittek, Rafael, Tom A.B. Snijders and Victor Nee (eds) (2013), *Handbook of Rational Choice Social Research*, Stanford, CA: Stanford University Press.
Ylikoski, Petri (2013), 'The (hopefully) last stand of the covering-law theory: a reply to Opp', *Social Science Information*, **53** (3), 383–93.
Ziegler, Rolf (1972), *Theorie und Modell [Theory and Model]*, München: Oldenbourg.
Ziegler, Rolf (2010), 'Deutschsprachige Netzwerkforschung [Research on social networks in German speaking countries]', in Christian Stegbauer and Roger Häussling (eds), *Handbuch Netzwerkforschung*, Wiesbaden: VS Verlag, pp. 39–53.

Werner Raub - 9781789906851

4. Analytical sociology and complexity research

Michael Mäs[1]

1. INTRODUCTION

> It is indeed certain that in the living cell there are only molecules of crude matter. But they are in association, and it is this association which is the cause of the new phenomena which characterise life, even the germ of which it is impossible to find in a single one of these associated elements. This is because the whole does not equal the sum of its parts; it is something different, whose properties differ from those displayed by the parts from which it is formed. […] By virtue of this principle, society is not the mere sum of individuals, but the system formed by their association represents a specific reality which has its own characteristics. Undoubtedly no collective entity can be produced if there are no individual consciousnesses: this is a necessary but not a sufficient condition. In addition, these consciousnesses must be associated and combined, but combined in a certain way. It is from this combination that social life arises and consequently it is this combination which explains it. By aggregating together, by interpenetrating, by fusing together, individuals give birth to a being, psychical if you will, but one which constitutes a psychical individuality of a new kind. (Durkheim 1982, pp. 128–9)

The above quote from 1895 by Emile Durkheim summarizes sociology's core contribution to the understanding of human behavior and justifies the existence of sociology as a distinct discipline. In a nutshell Durkheim argued that "sociology is not the appendage of any other science, it is itself a distinct and autonomous science" (Durkheim 1982, p. 162), because there is a nontrivial relation between phenomena on the macro-level and the behavior of individuals on the micro-level. Sometimes even a micro-theory that perfectly describes the behavior of individuals does not suffice to understand the macro-phenomena that arise from individual behavior. This apparent inconsistency between micro and macro-level is called the micro-macro-problem and is considered the core research problem of sociology. In particular, analytical sociologists argue that explaining and understanding macro-phenomena based on assumptions about individual behavior is sociology's core contribution to the understanding of human behavior (Schelling 1978; Coleman 1990; Hedström and Swedberg 1998; Udehn 2002).

Sociology is not the only discipline facing micro-macro problems, as the reference to molecules in Durkheim's quote suggests. In fact, micro-macro problems are studied in disciplines as diverse as physics, biology, mathematics, computer science and economics. A prominent methodological approach that many researchers in these disciplines follow is called "complexity research". Complexity research is defined as the study of systems that consist of multiple micro-entities that are interdependent in that they exert influence on each other. These interdependencies, which Durkheim referred to as "combinations" and "associations" in the quote, can generate unexpected macro-outcomes. Examples of complex systems and

[1] Acknowledgements: The arguments put forward in this chapter are inspired by discussions with my academic mentors, Karl-Dieter Opp, Michael Macy, Dirk Helbing and Andreas Flache. In particular, Dirk Helbing's input helped shape the ideas presented here.

Michael Mäs - 9781789906851

constituting micro entities are matter and particles; a school of fish and individual fish; an avalanche and ice crystals; the Internet and computers, a power grid and power plants.

The first aim of this chapter is to identify important commonalities between typical sociological research problems and phenomena studied by complexity researchers. I show that both fields focus on macro-phenomena that emerge from the behavior of interdependent micro entities (Sawyer 2005; Cherkaoui 2008). This commonality is abstract, but insights, methods and models developed by complexity researchers can nevertheless contribute to solving research puzzles in sociology. To be sure, I do not argue that human collectives behave according to the same principles as, for instance, molecules. Nevertheless, although theories of nonhuman complex systems and sociological theories are concerned with fundamentally different phenomena, their theoretical assumptions may translate into mathematically very similar models that can be studied with similar methods. When models are sufficiently similar, it is worthwhile to explore whether findings from other disciplines can be translated back into meaningful sociological insights.

Next I compare the complexity approach to prominent methodological approaches to sociology, showing that complexity research shares core principles with both structural and structural-individualistic approaches to sociology. However, in particular, the approach put forward by analytical sociologists overlaps on the crucial dimensions with complexity research (Hedström and Swedberg 1998; Hedström 2005; Hedström and Ylikoski 2010).

The second aim of this chapter is to identify aspects where sociological research can profit from the complexity perspective. While I conclude that analytical sociology shares central aims and principles with complexity science, I point to approaches and practices in analytical sociology that are puzzling from a complexity perspective and frankly formulate recommendations on how to refocus research on sociology's core contribution to the understanding of human collectives. In a nutshell, I argue that complexity research demonstrates that sociologists should avoid the so-called "representative-agent approach", focus more on the identification of social mechanisms, and strive more for general rather than so-called "middle-range theories". From the complexity perspective, further roadblocks to progress in sociology are: a too strong focus on random samples in empirical research; the fact that most theories are based on deterministic assumptions about human behavior; and that sociologists tend to think in individual-level variables—a method of theorizing that neglects the sociological dimension of social systems.

2. THE MICRO-MACRO PROBLEM IN SOCIOLOGY

While many sociologists agree with Durkheim's argument that the whole can be different from the sum of the parts, there is little agreement with one of the conclusions that Durkheim drew from this notion. Durkheim argued that if the whole is different from the sum of the parts, then theories explaining the whole should not refer to the parts. "The determining cause of a social fact must be sought among antecedent social facts, and not among the states of the individual consciousness" (Durkheim 1982, p. 134). Figure 4.1 illustrates this so-called "structural approach" to sociology, showing Turner's visualization of Durkheim's theory of social differentiation (Turner 1995). It shows that Durkheim used only macro-phenomena to explain social differentiation.

Durkheim's structural approach has been subject to important criticism. On the one hand there is no methodological reason to fundamentally reject macro explanations if they were validly

Michael Mäs - 9781789906851

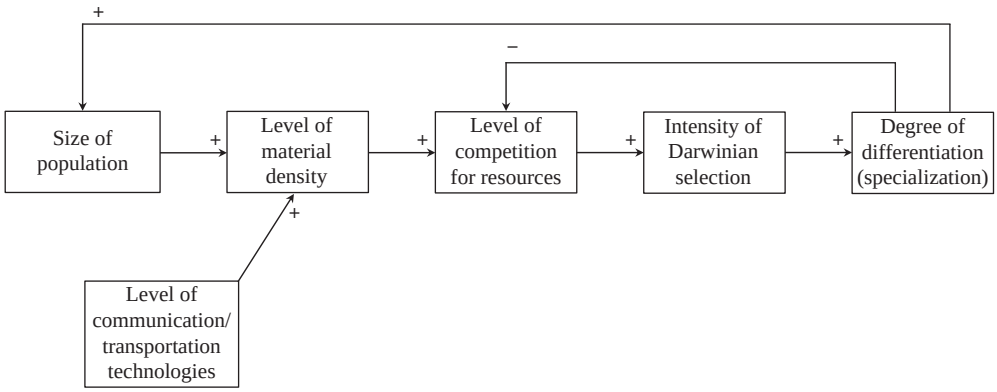


Figure 4.1 Summary of Durkheim's theory of social differentiation (taken from Turner 1995, p. 20)

derived according to the principles of the covering law model and its conditions of adequacy (Hempel and Oppenheim 1948; Popper 1959).[2] On the other hand it has been demonstrated that Durkheim's conclusion that macro-phenomena cannot be traced back to assumption of individual behavior is not true. Accordingly, many sociologists argue that macro-phenomena can and should be explained based on assumptions about individual behavior, a paradigm that is called the "structural-individualistic research program" (Homans 1958; Coleman 1990; Opp 2011; Raub, Buskens and van Assen 2011). The core idea to explain relationships between macro phenomena with assumptions about individual behavior has been introduced to sociology mainly by Homans (1958), but can be traced back to the Scottish moral philosophers (see Schneider 1967 for an overview) (please also see Raub's chapter in this book).

Figure 4.2 illustrates the structural-individualistic research program using one of the macro relationships from Durkheim's theory of social differentiation as an example.[3] It shows a so-called "Coleman-boat", a stylized scheme to represent structural-individualistic models (McClelland 1961; Coleman 1990; Raub, Buskens and van Assen 2011) that has turned into the work horse of the structural-individualistic research program. In contrast to the macro-model shown in Figure 4.1, the relationship between the level of competition for resources and the degree of differentiation in a population (see Arrow 1 in Figure 4.2) is represented by a dashed arrow to visualize that this relationship is not causal but results from relationships 2 through 4. Arrow 2 shows that according to this structural-individualistic model the level of competition for resources in a population creates an incentive for individuals to increase their uniqueness. For instance, when the number of bakeries in a town increases at some point in time bakers will experience increased competition. Arrow 2 suggests that each individual baker will have a stronger incentive to offer products that differ from those of their competitors. Arrow 3 adds that these incentives will motivate bakers to develop products that competitors do not offer.

[2] Important sociological macro theories do not satisfy the conditions of adequacy (Hempel 1994). However, this is not rooted in the fact that these theories study macro-phenomena. These are just poor theories.

[3] A less simplistic micro-macro model of Durkheim's macro-theory has been developed by Mäs et al., (2010) using the complexity approach.

Michael Mäs - 9781789906851

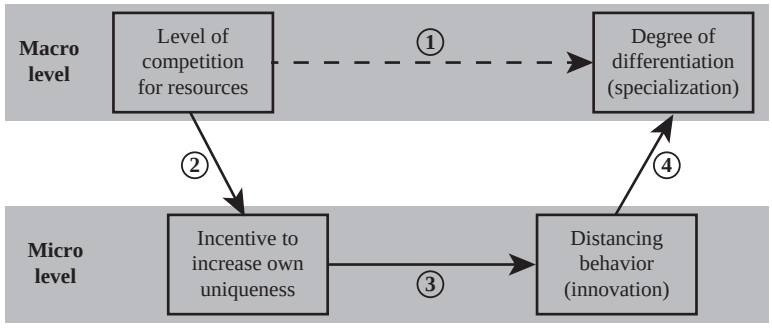


Figure 4.2 Example of a structural-individualistic explanation

Finally, Arrow 4 represents that the distancing behavior by individual bakers will translate to an increased degree of differentiation in the town. In other words, bakers will have diversified and customers will be able to choose from a bigger range of products.

In addition to the notion that macro-phenomena can be reduced to individual behavior, the example shown in Figure 4.2 illustrates two important differences to Durkheim's approach to sociology. First, in this example the whole is *not* different from the sum of the parts. In fact, if one defines the degree of differentiation amongst bakers as the number of different products available on the market and distancing behavior as the act of introducing a new product, then the whole is actually identical to the sum of the parts. Second, the example explains the macro relationship without reference to the interdependencies between individuals that Durkheim considered critical. In fact, theories like the one shown in Figure 4.2 have been criticized for adopting a so-called "representative-agent approach" (see e.g. Kirman 1992), which suggests that an accurate model of the behavior of an isolated individual suffices to explain macro-phenomena.[4] Durkheim, in contrast, argued that sociology as a discipline is concerned with phenomena where this assumption is not true, as the quote illustrates.

However, there are many structural-individualistic explanations in sociology where the whole differs from the sum of the parts and where interdependencies matter in a critical way (Epstein and Axtell 1996; Mäs, Flache and Helbing 2010; Mäs and Flache 2013; e.g. Mäs and Dijkstra 2014). The most prominent example is certainly Schelling's (1971) model of residential segregation (see also Sakoda 1971; Hegselmann 2017). With this model, Schelling famously demonstrated that cities can segregate into ethnically homogenous but mutually different neighborhoods even when all inhabitants are surprisingly tolerant, as Figure 4.3 illustrates. He represented a city as a grid with 50*50=2500 cells where each cell represents a home. Every actor belongs either to the black or the grey ethnic subgroup. White cells represent empty flats. At the outset (Panel a), actors are placed randomly on the grid, which creates a setting with minimal residential segregation. Next, Schelling assumed that actors sequentially

[4] Criticism of the representative-agent approach is based on two main arguments. First, it is often misleading to assume that all individuals have the same motives and beliefs, perceive the same restrictions, and decide based on the same decision rules. Second, it can be problematic to assume that having understood the decisions of a single member of a population suffices to predict the behavior of the collective. While complexity research is in line with both arguments against the representative-agent approach, I focus here on the second, since it closely corresponds to Durkheim's view on sociology.

Michael Mäs - 9781789906851

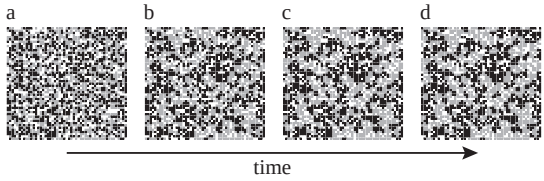


Figure 4.3 Typical dynamic of Schelling's model of residential segregation

decide whether they want to stay at their current home or move to an empty one. That is, one of the actors is picked at random and the share of neighbors of the same color on eight surrounding cells is calculated. The picked actor moves to a randomly selected empty flat when more than 70 per cent of her neighbors belong to the other ethnic group. This simple rule implements that actors are tolerant, as they accept to live in a neighborhood where their own group is the minority.

Panels b–d in Figure 4.3 show that even a city of very tolerant inhabitants will segregate. In Panel d, dynamics have reached a rest point where no actor occupies a flat where more than 70 per cent of the neighbors are of the opposite color. As a consequence actors do not move any longer. Strikingly, Panel d shows that the city is highly segregated, even though all actors are assumed to be rather tolerant.

Schelling's segregation model illustrates Durkheim's metaphor that the whole can be different from the sum of the parts, a notion that has been criticized for being "unusually ambiguous, metaphorical, and vague" (Nagel 1979, p. 381). The collective pattern of segregation (the whole) seems to be in conflict with the motives that drive individuals' moving decisions (the parts). Even though Schelling perfectly specified the decision rules that guide the behavior of all agents and no further assumptions were added to the model, it is not clear why the city segregated. In other words, even though the individual behavior is perfectly understood, it is not obvious why the collective experiences segregation.

Segregation emerges despite high tolerance because actors do not decide in isolation. They act interdependently. Whenever an actor moves, their ethnicity becomes less represented in their old neighborhood and more represented in their new neighborhood. This, in turn, has the potential to motivate further members of their own ethnicity to leave their old neighborhood and to motivate members of the other group to leave their new neighborhood. Both consequences are not intended by the actor, but make the old and the new neighborhood more homogenous and, thus, increase residential segregation. In a nutshell, the reason tolerant cities segregate is rooted in the interdependencies between actors, which closely corresponds to what Durkheim referred to as "associations" and "combinations". Thus, while the collective pattern of segregation can be generated (explained according to the covering law model) with Schelling's assumptions about individual behavior, even a perfect understanding of these micro assumptions is not sufficient to understand the behavior of the whole if the interdependencies between the parts are not understood. In this sense, the whole differs from the sum of the parts.

Within the structural-individualistic research program, the framework of analytical sociology directly addresses the macro-consequences of interdependencies between individuals (Hedström and Swedberg 1998; Hedström 2005; Hedström and Ylikoski 2010). Analytical sociologists call the described process that generates residential segregation the "mechanism" responsible for segregation and argue that identifying the mechanism responsible for the emergence of a

Michael Mäs - 9781789906851

macro phenomenon is a critical contribution to its explanation, although the mechanism is not explicitly included as an assumption in the theory. Residential segregation, for instance, can be adequately explained according to the principles of the covering law model without explicitly assuming the mechanism, as Schelling demonstrated. However, the mechanism is a logical consequence of the other assumptions and is, thus, hidden. Analytical sociologists consider theories unsatisfactory that hide an ingredient critical to understanding why the theory explains a given phenomenon (Hedström and Ylikoski 2010). Therefore, analytical sociologists insist that mechanisms need to be identified and consider this the central contribution of sociology.

The emphasis of the identification of mechanisms is fundamentally sociological, in that it closely corresponds to what Durkheim defined as the central contribution of sociology as a discipline. An ideal-typical psychological approach, in contrast, would explain segregation based on theories that point to conditions under which actors seek to live in segregated neighborhoods. For instance, social-identity theory (Tajfel and Turner 1986) specifies conditions under which individuals avoid members of outgroups. While analytical sociology does not debate that such individual motives exist, analytical sociology fundamentally differs from this psychological approach in that it draws to attention that cities can segregate even when individuals are tolerant and that segregation can be an unintended consequence of individual behavior. What is more, analytical sociology focuses on understanding how these unintended consequences arise, calling for the identification of mechanisms.

3. THE MICRO-MACRO PROBLEM IN COMPLEXITY RESEARCH

Complexity researchers study so-called "complex systems". By definition, a complex system has two crucial ingredients. First, a complex system consists of interdependent components. Second, the behavior of the system as a whole cannot be readily inferred from the behavior of an individual component (Bar-Yam 2003; Page 2015). In contrast, in complex-systems research macro-phenomena "emerge" from the behavior of the components. In complexity research a system-level phenomenon is called "emergent" when it is not possible to infer it from the behavior of isolated system components but arises from the interaction between multiple components. A typical example of a complex system is a termite mound, a colossal structure with an intricate ventilation and cooling system that was constructed by thousands of termites. Termite mounds are complex because even a perfect description of an individual termite and its behavior will not suffice to understand how the mound was constructed. In contrast, the termite mound "emerges" from the interaction of many termites.

In this section, I describe ideal-typical examples of complex systems, comparing complexity research with the different sociological approaches described above.

3.1 Example 1: The Chaos Game

The chaos game is a typical example of a mathematical complex system (Feldman 2012). It is based on very simple assumptions that generate an intriguing emergent outcome. The chaos game is based on the following five rules:

1. Draw an equilateral triangle on a canvas.
2. Pick a random spot inside the triangle and mark it.

Michael Mäs - 9781789906851

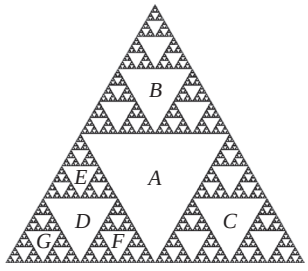


Figure 4.4 The Sierpiński triangle

3. Randomly pick one of the three corners.
4. From the previous mark, move half of the distance toward the picked corner and mark the new location.
5. Continue with step 3.

Before continuing reading, I invite the reader to speculate about the outcome of iterating many times through Steps 3–5. When I present the chaos game to students, for instance, many predict that the triangle will be completely filled with marks. Another frequent conjecture is that marks will converge in the center of the triangle. Both conjectures are false.

The pattern that the five rules generate is shown in Figure 4.4 and is called the "Sierpiński triangle", named after Wacław Sierpiński who described it for the first time in 1915. The Sierpiński triangle consists of four triangles, each of which consists of four triangles, each of which consists of four triangles, and so on. Every triangle will consist of four triangles no matter how deep one zooms in.

What mechanism generates the Sierpiński triangle? First, it is important to understand that all marks will be placed inside the triangle. This is caused by Rule 4, which says that the marks will always be placed between the previous mark and one of the three corners, which makes it impossible to end up outside the perimeter of the triangle. Second, it is impossible to place a subsequent mark inside the center triangle (Triangle A in Figure 4.4), because to place a mark inside the center triangle, the previous mark would need to be outside the main triangle, which is impossible. Furthermore, it is also impossible to place a mark inside triangles B, C and D, as this would require that the previous mark is inside triangle A, which we just excluded. Next, if it is impossible to place a mark inside triangle D, then also triangles E, F and G will never be marked, and so on.[5]

The chaos game has both ingredients of a complex system. First, Step 4 creates *interdependencies* between the components (the marks). Second, the pattern that the five rules generate is not explicitly built into the five rules. To be sure, the five rules do logically imply the Sierpiński triangle, as they generate the described mechanism. However, the rules that lead

[5] Strictly speaking, the exact shape of a Sierpiński triangle emerges only when the very first mark is placed on the black part of it. When the first mark is placed on a spot that is shown in white in Figure 4.4, then all subsequent marks will also be located on white areas. Nevertheless, the marks will converge on the Sierpiński triangle. The emerging pattern will, thus, be very similar to the one shown in Figure 4.4. To see this, suppose that the first mark happens to be placed inside Triangle A. The subsequent mark will be placed either inside B, C or D. However, it is excluded that any further marks will be inside Triangle A, because this would require that the respective previous mark is outside the main triangle. We have already established that this excludes that further marks are placed inside B, C, and D and so on.

Michael Mäs - 9781789906851

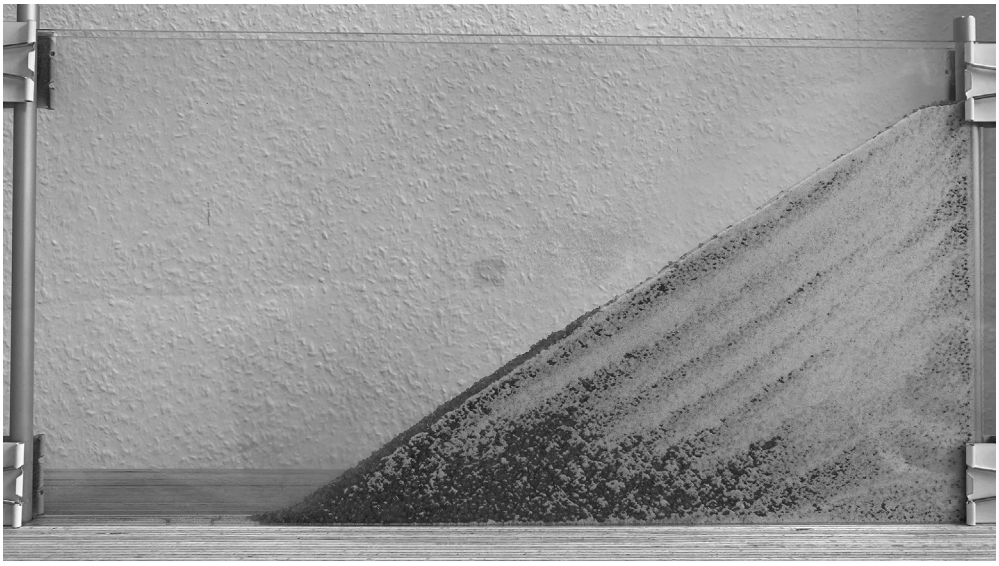


Figure 4.5 Spontaneous stratification (dark areas are red sugar; light areas are white sand)

to each individual mark do not make explicit that marks collectively create the Sierpiński triangle. Knowing the underlying mechanisms is critical to understanding why the Sierpiński triangle emerges.

3.2 Example 2: Spontaneous Stratification

This example stems from physics and can also be replicated with little effort (Makse et al. 1997). To prepare it get a piece of plywood, drill two holes into it, and stick a bar into each hole. Next, attach two panes of (acrylic) glass to the sticks, creating a gap between the panes. Finally, prepare an equal-volume mixture of half white sand and half red sugar.

Next, gently pour the mixture of sand and sugar into the gap between the two panes of glass and watch what happens. At the beginning, sand and sugar will form a well-mixed pile but after some time alternating layers of red sugar and white sand will emerge. This pattern is called "spontaneous stratification". Figure 4.5 shows a typical outcome.

Research suggests that the mechanism generating spontaneous stratification works as follows (Makse et al. 1997). When a pile has formed, grains will continue rolling down the pile. However, when they reach the base of the pile, the slope of the pile decreases, forming a kink at which grains are stopped. However, sand consists of small rounded grains, while sugar grains are much larger and cubical. As a consequence, the sugar grains role down further on top of the small grains of sand. Sand grains, in contrast, get caught in the spaces between sugar grains. This results in the formation of two layers with the sand underneath the sugar. Furthermore, as the flow of grains continues, the kink moves upward until it reaches the top of the pile and the cycle is repeated, forming another pair of layers.

The experiment is a typical example of a complex system. The main components of the system are the grains of sand and sugar. Each grain behaves according to the laws of physics,

Michael Mäs - 9781789906851

but grains also interact, affecting each other's positions. Second, the rules of physics that govern the behavior of each individual grain do not specify that the macro-pattern of spontaneous stratification emerges.

3.3 Conclusion

Table 4.1 provides further examples of complex systems from various disciplines. Each of the examples illustrates the core building blocks of a complex system: (i) interdependent components and (ii) emergence. There is also an increasing number of introductions to the study of complex systems in mathematics (Feldman 2012), physics (Nicolis and Prigogine 1989), biology (Camazine et al. 2001), and complexity research in general (Ball 2004; Page 2010).

There are evident commonalities between complexity research and the sociological approaches sketched in Section 2, as Table 4.2 summarizes. First, complexity research resembles the classic sociological view on human collectives that Durkheim put forward. In fact, Durkheim's quote mentions both building blocks of a complex system ("consciousnesses must be associated"; "the whole does not equal the sum of its parts"). Durkheim even argued that these building blocks justify sociology's status as a distinct discipline. The critical difference between Durkheim's approach to sociology and complexity research, however, is that Durkheim rejected the notion that macro-phenomena can be explained based on assumptions about the behavior of the components, which I call the "reducibility assumption" in Table 4.2.

Table 4.1 Examples of complex systems studied in various disciplines

Discipline	System	Components	Emergent property	Further information
Mathematics	Set of marks in a triangle	Marks	Sierpiński triangle	Movie,[1] (Feldman 2012)
Physics	Mixture of sand and sugar	Grains of sand and sugar	Spontaneous stratification	Movie,[2] (Makse et al. 1997)
Biology	Flock of birds	Birds	Murmuration	(King and Sumpter 2012)
Biology	Colony of bacteria	Bacteria	Clustering	(Camazine et al. 2001)
Biology	Population of fish	Fish	Fish schooling	(Hemelrijk and Hildenbrandt 2012)
Biology	Swarm of fire flies	Fire flies	Synchronized flashing	(Strogatz and Stewart 1993)
Pedestrian Dynamics	Busy corridor in a building	Pedestrians	Lane formation	(Helbing and Molnár 1995)
Engineering	Power grids	Power plants	Black outs	(Dobson et al. 2007)

Notes:
[1] https://www.youtube.com/watch?v=DcvY_45XFU4.
[2] https://www.simonsfoundation.org/multimedia/mathematical-impressions-multimedia/mathematical-impressions-spontaneous-stratification/.

Michael Mäs - 9781789906851

Table 4.2 Overview over approaches to sociology and their relations to complexity science

	Acknowledges the critical role of interdependencies	Focuses on phenomena where "the whole is more than the sum of parts"	Accepts reducibility assumption
Durkheim's structural approach	yes	yes	no
Representative agent approach	no	no	yes
Analytical sociology	yes	yes	yes
Complexity science	yes	yes	yes

Like complexity research, the structural-individualistic research program disagrees with Durkheim on the reducibility assumption, insisting that the reduction of macro-phenomena to micro-level behavior is possible. However, those contributions that follow the representative-agent approach within the structural-individualistic research program differ fundamentally from complexity research in that they ignore the two ingredients of complex systems. While the representative-agent approach may not exclude that individual behavior is interdependent and that interdependency can lead to unintended consequences, this approach to sociology does not direct attention to these central building blocks of a complex system.

The framework of analytical sociology, however, has a strong focus on emergent sociological phenomena and explanations based on interdependencies (Manzo 2014). Therefore, I conclude that there is no fundamental methodological difference between analytical sociology and complexity research.

On a more general notice, I conclude that complexity science and the discipline of sociology share features that leading representatives of each field have identified as the basic building blocks of their discipline. Both fields study the emergence of macro phenomena arising from the interdependencies between micro-entities. Thus, while I agree with Scott Page's conclusion that "a meaningful part of sociology overlaps with a substantial subset of the complexity sciences" (2015, p. 22), I add that the two fields actually overlap on their defining features.

4. A COMPLEXITY PERSPECTIVE ON ANALYTICAL SOCIOLOGY

The comparison between alternative methodological approaches to sociology and complexity science led me to the conclusion that there is no fundamental methodological difference between analytical sociology and complexity research. In this section, however, I point to differences between the work of many (analytical) sociologists and complexity researchers, arguing that insights from complexity science have the potential to foster scientific advances in sociology. I frankly formulate six recommendations, arguing that a complexity perspective will help sociologists refocus their research on sociology's original contribution to the understanding of human behavior, the study of emergent phenomena arising from social interdependencies.

Michael Mäs - 9781789906851

To be sure, the following is based on generalizations that fail to reflect the diversity of theoretical and empirical methods applied by analytical sociologists. There are many contributions by analytical sociologists that are very much in line with the recommendations I formulate. Nevertheless, there are also broadly accepted research practices in the field of analytical sociology that seem counter-productive from a complexity perspective.

4.1 Drop the Representative-Agent Approach! Avoid the Coleman Boat!

From the perspective of complexity research, the prominence of the representative-agent approach in contemporary sociological research and in parts of analytical sociology is puzzling. While the representative-agent approach shares with complexity research the emphasis of micro theories, it fails to address interdependencies between individuals and the unintended macro-consequences that arise from them. In contrast, according to both Durkheim and complexity research, sociology's key contribution to the understanding of human behavior is that interaction can give rise to phenomena that cannot be understood with a model of isolated individuals.

To be sure, to develop and test theories of individual behavior, it is often very useful to study individuals in isolation. However, as soon as one seeks to derive macro-predictions from a micro theory, the assumption that humans act in isolation is not only false but also potentially misleading. The assumption that individuals act in isolation is, thus, not an innocent assumption.

For the same reason, complexity researchers are puzzled by analytical sociologists' frequent use of the Coleman boat (e.g. Hedström and Ylikoski 2010). While there is no doubt that the Coleman boat allows one to illustrate the basic principles behind the structural-individualistic research program, it fails to visualize the potentially critical impact of the social interdependencies between actors. For instance, the Coleman boat fails to illustrate the described mechanism generating residential segregation in tolerant populations. While it describes all model assumptions about how actors perceive their environment and how they decide whether to move or not, the Coleman boat does not show how these assumptions translate to segregation. The Coleman boat, in other words, fails to visualize interaction and emergence, the two building blocks of complex systems and the core contributions of sociology. It omits the sociological perspective on human behavior, suggesting that a psychological model of social action suffices to explain the behavior of human collectives, a shortcoming that Coleman had already identified (Coleman 1986, 1990).

4.2 Identify and Study Mechanisms!

In the past decades, the focus of empirical social-scientific research has shifted very much to the micro level. Inspired by advances in economics and psychology, social scientists have elaborated models of individual behavior, studying for instance assumptions about pro-social preferences (Fehr and Gächter 2000, 2002), framing processes (Lindenberg and Frey 1993), and decision-making based on heuristics (Gigerenzer et al. 1999). Likewise, in the field of social-network analysis, new and powerful statistical methods have led to many contributions for testing hypotheses about how individuals form social relationships (Snijders et al. 2010).

On the one hand complexity researchers applaud research testing and elaborating micro models. True micro-macro explanations require true micro-models. On the other hand a big

Michael Mäs - 9781789906851

part of sociological research in this direction seems to be motivated by the aim to develop increasingly realistic micro-models rather than by sociological relevance (Hedström and Ylikoski 2011, 2014; Conte and Giardini 2016; Jager 2017).

The problem is that realistic micro-models tend to be very complex in terms of the number of assumptions they make about how individuals behave. However, the complexity that sociology is concerned with results from the interaction *between* individuals and not from the complexity *within* each individual. To be able to understand the complexity arising from interdependencies, sociologists should keep their micro-models as simple as possible. Thus, from the perspective of complexity research, the challenge is to develop micro-models that are sufficiently simple to allow the researcher to understand their macro-consequences and, at the same time, sufficiently complex to be meaningful.

Suppose, for example, you want to explain the macro-phenomenon of residential segregation based on a realistic micro-model of individuals' housing preferences and moving restrictions. This model would certainly take into account individuals' budget constraints; local variation in housing prices; the fact that many families choose neighborhoods based on characteristics of local schools; and geospatial aspects such as workplace proximity. A realistic model would also mimic individuals' decision-making processes as much as possible. While these aspects certainly contribute to the explanation of moving patterns on the micro-level, including them in a micro-macro model makes it more difficult to identify the described mechanism that is responsible for the emergence of segregation, because the decision of each modeled individual is more complicated.

From the complexity perspective, sociological research should be guided by the sociological mechanisms that generate unintended macro-consequences. That is, sociological research should depart from the analysis of the mechanisms and identify those aspects of individual behavior that put into motion the mechanism responsible for the systems' complexity. Empirical research should test hypotheses about the conditions under which these aspects affect individual decisions. Other aspects might be relevant from an empirical perspective and are therefore also important but they do not contribute to understanding the emergence of macro phenomena.

4.3 Reconsider the Focus on Random Samples!

From the complexity perspective, it is puzzling that random sampling is the dominant method of data gathering in sociology. The problem is that one of the critical assumptions underlying the statistical analysis of random samples is that observations are independent from each other. This assumes away that individuals exert influence on each other and that these interdependencies matter. This assumes away sociology.

Figure 4.6 illustrates this disadvantage of random samples. The graph on the left-hand side shows an artificial social network. Node colors represent, for instance, the political orientation of nodes (grey: conservative; black: liberal). Besides a few exceptions most nodes have the political orientation that is dominant in their ego-network, suggesting that social influence played an important role in the formation of the actors' political orientation. The graph on the right-hand side, however, shows the information that can be gathered with a random sample. While this random sample does allow the researcher to draw statistical inferences about the distribution of conservative and liberal orientations in the whole network, it does not allow one to test hypotheses about potential social-influence processes that affected actors' orientations.

Michael Mäs - 9781789906851

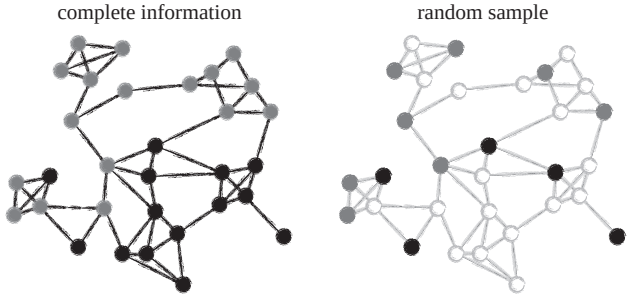


Figure 4.6 Disadvantage of random samples

To be sure, random samples provide information that can be critical to answering important research questions. They allow drawing statistical inferences about the entire population, providing a powerful tool do describe societies and collective trends. As a consequence, research studying important sociological phenomena such as inequality, segregation, discrimination and poverty rely on random sampling methods. Furthermore, random samples allow to test hypotheses about within-individual processes. For instance, one can test whether participants' political orientation correlates with their income, or their education. What is more, it is possible to statistically control for parts of the interdependencies between individuals (Snijders and Bosker 1999) and, sometimes, to measure social influences also in studies based on random samples (e.g. Opp and Gern 1993).

However, from the perspective of complexity science it is surprising that random sampling is the dominant approach in quantitative, empirical sociology, a field that is built on the notion that interdependencies matter. The dominant role of random sampling in sociology implies that sociologists might systematically overlook inherently sociological aspects contributing to social phenomena. James Coleman criticized already in 1986 that empirical research in sociology had shifted "from social processes within the community shaping the system's behavior to psychological or demographic processes shaping individual behavior" (Coleman 1986, p. 1315). He reasoned that advances in statistical methods had made random samples increasingly powerful tools and that "there was no comparable development of tools for analysis of the behavior of interacting systems of individuals or for capturing the interdependencies of individual actions as they combine to produce a system-level outcome". Advances in the statistical analysis of social-network data (Robins et al. 2007; Snijders et al. 2010) in tandem with new technical opportunities to measure social networks at a large scale (Lazer et al. 2009; Golder and Macy 2014) have the potential to fill this gap.

4.4 Do Not Only Think in Individual-Level Variables!

Many sociologists think in variables. That is, they develop a set of concepts (variables) describing entities on some level of analysis (e.g. individuals, organizations, societies) and study the relationships between these concepts (Lazarsfeld and Menzel 1969). The path diagram shown in Figure 4.1 is an example of such a variable analysis on the macro-level. It assumes that all societies can be described by six different characteristics and proposes causal relationships between these variables.

Michael Mäs - 9781789906851

Thinking in variables is a powerful method, but it fails to capture an important sociological dimension of social systems if it is applied only to variables on the level of individuals or collectives. On the one hand, thinking in variables allows one to represent complex theories in a very structured and intuitive way, as Figure 4.1 illustrates. A second strength is that theoretical models derived with variable thinking can be translated directly into statistical models, which makes it straightforward to test variable models. On the other hand, thinking in variables is mainly used to specify relationships between attributes of individuals or collectives, and not between the units of analysis. This neglects the interdependencies between the units of analysis and thus overlooks the sociological dimension of social systems (Macy and Willer 2002; Page 2015).

The weakness of variable-based thinking is illustrated by DellaPosta et al.'s (2015) discussion of why liberals drink lattes. Thinking in variables one can explain the correlation between the variables "political orientation" (A) and "coffee preference" (B) in three ways. First, political orientation might have a causal effect on coffee preferences (A→B). Second, coffee preferences might causally affect individuals' political orientation (B→A). Third, there might be a third variable, C, that affects both A and B, creating a spurious correlation between the two variables. Each of the three explanations points to a within-individual process. DellaPosta et al., however, point to a fourth, surprisingly simple explanation that cannot be derived by thinking in individual-level variables. With a formal model, they demonstrate that a correlation between A and B can also emerge from the interplay of homophily and social influence, processes that can reinforce even small and random differences between groups. Suppose that coffee preferences were completely random and causally unrelated to political orientation. Randomness, however, will always generate small fluctuations, making it possible that at some moment there are a few more liberals than conservatives who drink lattes in a sub-community of the population. Even a tiny and statistically insignificant relationship between the two variables can be reinforced if, in this community, liberals tend to consume their coffee together with other liberals (homophily) and exert influence on their interaction partners. Social influence will, thus, convince more liberals to drink lattes, intensifying the correlation between political orientation and coffee preferences (Mäs and Flache 2013; Mäs et al. 2013).

Variable-based thinking applied to individuals fails to capture this fourth candidate explanation, because it focuses on within-individual processes, creating "a 'blind spot' for the possibility that respondents' opinions were influenced by their peers" (DellaPosta et al. 2015, p. 1486) and limiting the researcher to psychological explanations. To be sure, it is an empirical question whether the explanation that DellaPosta et al. proposed is true. Furthermore, hypotheses following from this fourth explanation can also be translated into variable statements, which helps testing them empirically. For instance, the explanation proposed by DellaPosta et al. implies that the variable "preference for lattes" is correlated with the number of friends who also prefer lattes. Furthermore, variable-based thinking does capture influence between individuals when it is applied to pairs of individuals rather than only individuals (Hummell and Sodeur 1997). That is, one can use network ties as the unit of analysis and describe these ties with two variables capturing the coffee preferences of each network node. A causal relationship between these variables would represent social influence. Thus, variable-based thinking can be a very powerful approach to research on the relationships between individuals. However, considering that sociology's unique contribution to the understanding of human behavior is the study of collective patterns arising from interdependencies,

Michael Mäs - 9781789906851

complexity researchers are puzzled by how many hypotheses in the contemporary sociological literature have been derived by thinking in variables describing individuals, rather than from an analysis of the interdependencies between individuals.

4.5 Strive for General Theories!

Sociology has brought about many general theories (Turner 1974, 1995), but unlike in many other disciplines, these efforts have hardly contributed to the development of a sociological canon, a set of widely accepted assumptions, predictions and methods. The core weakness of sociology's existing grand theories is that they are too remote from reality and, as a consequence, hard if not impossible to test empirically (Merton 1957).

Having made these experiences, many analytical sociologists advocate a so-called "middle-range approach" (Merton 1957; Boudon 1981; Hedström and Udehn 2009; Hedström and Ylikoski 2010). That is, rather than developing grand theories applicable to various sociological contexts and phenomena, many sociologists develop theories with a limited scope. Sociologists then test middle-range theories in their specific empirical context, hoping to develop multiple accepted models for different contexts that can later be generalized into an overarching, general theory. For instance, sociology has rich literatures on the emergence of social institutions, on their stability and evolution, and on their decay (Ellickson 1991; Young 1993; Greif and Laitin 2004; e.g. Acemoglu and Robinson 2012). However, there is no general model of institutions that captures all of these phenomena.

Many complexity researchers are puzzled by the middle-range approach, coming from disciplines where the generality of a theory is considered a central quality criterion. Physicists, for instance, have a single model of matter, rather than separate theories of solids, liquids and gases. The central advantage of this general model is that it does not only describe all three states, but that it also specifies when and under which conditions matter moves from one state to the other. What is more, this model also helped discover states of matter that were unknown before (plasma, Bose–Einstein condensates) because they are hardly observed on earth.

The middle-range approach has a major shortcoming: there is no method of inducing general theories from a set of middle-range theories, a problem that is rooted in Hume's famous problem of induction (Hume 1955; Popper 1959). To be sure, it is possible to generalize theories. In fact, the history of physics as a discipline is a sequence of generalizations. The problem, however, is that there is no scientific method that guides induction, which is the core weakness compared to the opposite approach, deduction (for a more detailed discussion see Opp 2013).

Despite sociology's problematic history with grand theories, advances in other disciplines that study micro-macro problems should encourage sociologists to strive for general theories and suggest that sociological research should be guided by this motive. In addition to applying theories to very specific cases, sociologists should be bolder and explore the boundaries of their theories' scope, testing whether their theories can be abstracted and usefully applied to other contexts than initially intended. To this end theoretical contributions should make explicit the scope of the proposed model, as well as discuss which phenomena are excluded and why. Next, empirical research should not only test predictions within the theory's scope, but also explore boundaries of the scope. For instance, it should be tested empirically whether or not middle-range theories of crime also apply to deviations from informal norms and conventions.

Michael Mäs - 9781789906851

4.6 Consider Randomness on the Micro-Level!

Most sociological theories are based on deterministic assumptions. That is, most micro-macro explanations assume that individual behavior is "completely determined, flawlessly executed, entirely knowable, and perfectly predictable" (Macy and Tsvetkova 2013, p. 324), which is not only unrealistic but also potentially misleading, as deviations from the assumptions of micro-models can have critical effects on macro-predictions, even when these deviations are rare and random (Macy and Tsvetkova 2013; Mäs and Helbing 2020). To be sure, very few social scientists would argue that human behavior follows deterministic patterns. Nevertheless, most theorizing abstracts from the random deviations from these patterns.

In other disciplines concerned with micro-macro problems randomness is a crucial building block of theoretical models. In physics, for instance, the concept of temperature is defined as the random microscopic motion of particles and is a key ingredient of models of turbulent fluid flows and the kinetic theory of gases (Nicolis and Prigogine 1977). Randomness also plays a critical role in biological theories of the evolution of species, where random mutations in the genes of individuals are a primary source of variation that allows species to adapt to changes in their environment.

The potentially critical effects of random deviations in social systems can also be illustrated with Schelling's segregation model. Schelling's original model is deterministic in that it assumes that individuals are either satisfied with their neighborhood or not. Furthermore, if and only if actors are dissatisfied, they will move to another place. As sketched above, segregation emerges because whenever an actor moves, they might make one of their old or new neighbors dissatisfied, potentially sparking cascades of further movement. Strikingly it turns out that segregation is intensified when Schelling's original micro-assumptions are relaxed by assuming that there is always a tiny probability that a satisfied actor moves (van de Rijt, Siegel and Macy 2009). This fosters segregation, because every moving has the potential to reactivate the social mechanism, sparking a new cascade that further intensifies segregation. Thus, while the deterministic model reaches a state of equilibrium at some point, deviations keep the segregation process running, even though deviations are rare and random.

Schelling's segregation model is not the only sociological theory where randomness on the micro-level affect macro-outcomes. Similar effects of randomness have been shown with theories of social norms (Young 2015), conventions (Young 1993), the dissemination of culture (Klemm et al. 2003; Huckfeldt et al. 2004), opinion polarization (Pineda et al. 2009; Mäs et al. 2010; Kurahashi-Nakamura et al. 2016), collective behavior (Granovetter 1978), and the production of collective goods (Kollock 1993). What is more, there is also empirical evidence supporting the hypothesis that deviations from otherwise dominant patterns of individual behavior can change macro-outcomes and that it can be predicted when deviations matter and when not (Mäs and Helbing 2020).

Despite these findings, sociologists seem to have an ambivalent relationship with the concept of randomness. One the one hand, randomness plays an important role in statistical models where typically more than half of the dependent variables' variance is unexplained and modelled as a random error term. On the other hand, randomness is often perceived as a residual component that needs to be stripped away in order to reveal systemic behavior. This may be true when individuals act in isolation. However, when individual behavior is interdependent, even random and rare behavior of an actor can affect the decisions of others,

Michael Mäs - 9781789906851

potentially sparking cascades that affect macro-outcomes. As interdependencies are an Integral part of sociological theories, it is important to make probabilistic rather than deterministic assumptions about individual behavior.

5. SUMMARY

It is an exciting time to be a sociologist. Sociological phenomena such as inequality, poverty, cohesion, social order, integration and cultural differentiation are increasingly recognized as major challenges for societies. As a result of the success of the Internet and rapid globalization, social systems are growing increasingly complex and interconnected, making it more important and more challenging to explain and predict the outcomes of societal dynamics.

Already Emile Durkheim, sociology's founding father, argued that sociology is concerned with collective phenomena arising from the interdependencies between individuals, proposing that these interdependencies can be responsible for complex, unintended and puzzling macro phenomena. Contributions to disciplines as diverse as economics, biology, computer science, physics and engineering demonstrate that complexity research provides theoretical and empirical tools to study interdependencies on the micro level and their complex macro consequences. In fact, my comparison of complexity research with prominent sociological approaches led to the conclusion that complexity research shares core principles with prominent methodological approaches to sociology. However, in particular the framework of analytical sociology is based on congruent methodological principles.

Obviously there are limits to the application of methods and theoretical concepts from other disciplines to sociological questions. Exploring these limits while exploiting the enormous potential of interdisciplinary research is a core challenge for future sociological research.

REFERENCES

Acemoglu, D. and J.A. Robinson (2012), *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. NY: Crown Publishers.

Ball, P. (2004), *Critical Mass: How One Thing Leads to Another*. New York: Heinemann/Farrar, Straus & Giroux.

Bar-Yam, Y. (2003), *Dynamics of Complex Systems*. Boulder (Colorado): Westview Press.

Boudon, R. (1981), *The Logic of Social Action. An Introduction to Sociological Analysis*. London: Routledge and Kegan Paul.

Camazine, S. et al. (2001), *Self-Organization in Biological Systems*. Princeton, NJ: Princeton University Press.

Cherkaoui, M. (2008), *Durkheim and the Puzzle of Social Complexity*. Oxford: Bardwell Press.

Coleman, J.S. (1986), "Social theory, social research, and a theory of action", *American Journal of Sociology*, **91** (6), 1309–35.

Coleman, J.S. (1990), *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

Conte, R. and F. Giardini (2016), "Towards computational and behavioral social science", *European Psychologist*. doi: 10.1027/1016-9040/a000257.

DellaPosta, D., Y. Shi and M. Macy (2015), "Why do liberals drink lattes?" *American Journal of Sociology*, **120** (5), 1473–511. doi: 10.1086/681254.

Dobson, I. et al. (2007), "Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization", *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **17**, 026103. doi: 10.1063/1.2737822.

Durkheim, E. (1982), *The Rules of Sociological Method*, edited by S. Lukes. New York: The Free Press.

Ellickson, R.C. (1991), *Order without Law*. Cambridge, MA: Harvard University Press.

Epstein, J.M. and Axtell, R.L. (1996), *Growing Artificial Societies: Social Science from the Bottom Up*. Washington DC: The Brookings Institution.

Fehr, E. and S. Gächter (2000), "Cooperation and punishment in public goods experiments", *American Economic Review*, **90** (4), 980–94.

Michael Mäs - 9781789906851

Fehr, E. and S. Gächter (2002), "Altruistic punishment in humans", *Nature*, **415** (6868), 137–40.
Feldman, D.P. (2012), *Chaos and Fractals: An Elementary Introduction*. Oxford: Oxford University Press.
Gigerenzer, G., P.M. Todd and the ABC Research Group (1999), "Fast and frugal heuristics: The adaptive toolbox", in G. Gigerenzer, P.M. Todd and the ABC Research Group, (eds), *Simple Heuristics That Make Us Smart*. New York and Oxford: Oxford University Press, pp. 3–34.
Golder, S.A. and M.W. Macy (2014), "Digital footprints: Opportunities and challenges for online social research", *Annual Review of Sociology*, **40** (1), 129–52. doi: 10.1146/annurev-soc-071913-043145.
Granovetter, M. (1978), "Threshold models of collective behavior", *American Journal of Sociology*, **83** (6), 1420–43. doi: 10.1086/226707.
Greif, A. and D.D. Laitin (2004), "A theory of endogenous institutional change", *American Political Science Review*, **98** (4), 633–52.
Hedström, P. (2005), *Dissecting the Social: On the Principles of Analytical Sociology.* Cambridge: Cambridge University Press.
Hedström, P. and R. Swedberg (1998), *Social Mechanisms. An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
Hedström, P. and P. Ylikoski (2010), "Causal mechanisms in the social sciences", *Annual Review of Sociology*, **36** (1), 49–67. doi: 10.1146/annurev.soc.012809.102632.
Hedström, P. and P. Ylikoski (2011), "Analytical sociology", in *The SAGE Handbook of the Philosophy of Social Sciences*. London: SAGE, pp. 386–98. doi: 10.4135/9781473913868.n19.
Hedström, P. and P. Ylikoski (2014), "Analytical sociology and rational-choice theory", in G. Manzo (ed.), *Analytical Sociology*. Chichester, UK: John Wiley & Sons, pp. 53–70. doi: 10.1002/9781118762707.ch02.
Hedström, P. and L. Udehn (2009), "Analytical sociology and theories of the middle range", in P. Hedström and P. Bearman (eds), *The Oxford Handbook*. Oxford: Oxford University Press, pp. 25–47.
Hegselmann, R. (2017), "Thomas C. Schelling and James M. Sakoda: The intellectual, Technical, and social history of a model", *Journal of Artificial Societies and Social Simulation*, JASSS, **20** (3), 15. doi: 10.18564/jasss.3511.
Helbing, D. and P. Molnár (1995), "Social force model for pedestrian dynamics", *Physical Review*, *E*, **51** (5), 4282–6.
Hemelrijk, C.K. and H. Hildenbrandt (2012), "Schools of fish and flocks of birds: Their shape and internal structure by self-organization", *Interface focus*, *The Royal Society*, **2** (6), 726–37. doi: 10.1098/rsfs.2012.0025.
Hempel, C.G. (1994), "The Logic of Functional Analysis", in M. Martin and L.C. McIntyre (eds), *Readings in the Philosophy of Social Science*. MIT Press, pp. 349–75.
Hempel, C.G. and P. Oppenheim (1948), "Studies in the logic of explanation", *Philosophy of Science*, **15**, 135–75.
Homans, G.C. (1958), "Social behavior as exchange", *American Journal of Sociology*, **63** (3), 597–606.
Huckfeldt, R., P.E. Johnson and J. Sprague, J. (2004), *Political Disagreement. The Survival of Diverse Opinions within Communication Networks*, Cambridge: Cambridge University Press.
Hume, D. (1955), *An Enquiry Concerning Human Understanding*, Indianapolis: Bobbs-Merrill.
Hummell, H.J. and W. Sodeur (1997), "Structural analysis of social networks with respect to different levels of aggregation", *Mathématiques et sciences humaines*, **137**, 37–60.
Jager, W. (2017), "Enhancing the realism of simulation (EROS): On implementing and developing psychological theory in social simulation", *Journal of Artificial Societies and Social Simulation* **20** (3), 14. doi: 10.18564/jasss.3522.
King, A. and D. Sumpter (2012), "Murmurations", *Current Biology*, **22** (4), 112–4.
Kirman, A.P. (1992), "Whom or what does the representative individual represent?", *The Journal of Economic Perspectives*, **6** (2), 117–36.
Klemm, K. et al. (2003), "Global culture: A noise-induced transition in finite systems", *Physical Review E*, **67** (4), 045101(R).
Kollock, P. (1993), "'An eye for an eye leaves everyone blind': Cooperation and accounting systems", *American Sociological Review*, **58**, 768–86.
Kurahashi-Nakamura, T., M. Mäs and J. Lorenz (2016), "Robust clustering in generalized bounded confidence models", *Journal of Artificial Societies and Social Simulation*, **19** (4), 7.
Lazarsfeld, P.F. and H. Menzel (1969), "On the relation between individual and collective properties", in A. Etzioni (ed.), *Complex Organizations. A Sociological Reader*. New York: Holt, Rinehart and Winston, pp. 422–40.
Lazer, D. et al. (2009), "Computational social science", *Science*, **323** (5915), 721–3.
Lindenberg, S. and B.S. Frey (1993), "Alternatives, frames, and relative process: A broader view of rational choice theory", *Acta Sociologica*, **36** (3), 191–205.
Macy, M. and M. Tsvetkova (2013), "The signal importance of noise", *Sociological Methods & Research*, **44** (2), 306–28. doi: 10.1177/0049124113508093.
Macy, M.W. and R. Willer (2002), "From factors to actors: Computational sociology and agent-based modeling", *Annual Review of Sociology*, **28** (1), 143–66.
Makse, H.A., S. Havlin, P.R. King and H.E. Stanley (1997), "Spontaneous stratification in granular mixtures", *Nature*, **386** (27), 379–82.

Michael Mäs - 9781789906851

Manzo, G. (2014), "Problem shift in sociology: Mechanisms, generic instruments, and fractals", in G. Manzo (ed.), *Analytical Sociology*: *Actions and Networks*, Chichester, UK: John Wiley & Sons, pp. 419–26. doi: 10.1002/9781118762707.ch17.
Mäs, M., A. Flache, K. Takács and K.A. Jehn (2013), "In the short term we divide, in the long term we unite: Demographic crisscrossing and the effects of faultlines on subgroup polarization", *Organization Science*, **24** (3), 716–36.
Mäs, M. and A. Flache (2013), "Differentiation without distancing. Explaining opinion bi-polarization without assuming negative influence", *Plos One*, **8** (11).
Mäs, M., A. Flache and D. Helbing (2010), "Individualization as driving force of clustering phenomena in humans", *PLoS Computational Biology*, **6** (10), e1000959.
Mäs, M. and D. Helbing (2020), "Random deviations improve micro-macro predictions. An empirical test", *Sociological Methods & Research*, **49** (2), 387–417.
Mäs, M. and J. Dijkstra (2014), "Do intergroup conflicts necessarily result from outgroup hate?", *Plos One*, **9** (6).
McClelland, D. (1961), *The Achieving Society*. Glencoe, IL: The Free Press.
Merton, R.K. (1957), *Social Theory and Social Structure*. Glencoe, IL: The Free Press.
Nagel, E. (1961), *The Structure of Science: Problems in the Logic of Scientific Explanation.* San Diego: Harcourt, Brace & World
Nicolis, G. and I. Prigogine (1977), *Self-Organizationin Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations.* New York: Wiley Interscience.
Nicolis, G. and I. Prigogine (1989), *Exploring Complexity: An Introduction*. New York: W.H. Freeman.
Opp, K-D. (2011), "Modeling micro-macro relationships: Problems and solutions", *The Journal of Mathematical Sociology*, Taylor & Francis Online, **35** (1–3), 209–234. doi: 10.1080/0022250X.2010.532257.
Opp, K-D. (2013), "What is analytical sociology? Strengths and weaknesses of a new sociological research program", *Social Science Information*, 329–60. doi: 10.1177/0539018413483939.
Opp, K-D. and C. Gern (1993), "Dissident groups, personal networks, and spontaneous cooperation: The East German Revolution of 1989", *American Sociological Review*, **58** (5), 659–80.
Page, S.E. (2010), *Diversity and Complexity*. Princeton, NJ: Princeton University Press.
Page, S.E. (2015), "What sociologists should know about complexity", *Annual Review of Sociology*, *Annual Reviews*, **41** (1), 21–41. doi: 10.1146/annurev-soc-073014-112230.
Pineda, M., R. Toral and E. Hernandez-Garcıa (2009), "Noisy continuous-opinion dynamics", *Journal of Statistical Mechanics: Theory and Experiment*, P08001.
Popper, K.R. (1959), *The Logic of Scientific Discovery*. London: Hutchinson.
Raub, W., V. Buskens and M.A.L. van Assen (2011), "Micro-macro links and microfoundations in sociology", *The Journal of Mathematical Sociology*, **35** (1–3), 1–25. doi: 10.1080/0022250X.2010.532263.
van de Rijt, A., D. Siegel and M. Macy (2009), "Neighborhood chance and neighborhood change: A comment on Bruch and Mare", *American Journal of Sociology*, **114** (4), 1166–80.
Robins, G. et al. (2007), "An introduction to exponential random graph (p*) models for social networks", *Social Networks*, **29** (2), 173–91.
Sakoda, J.M. (1971), "The checkerboard model of social interaction", *The Journal of Mathematical Sociology*, **1** (1), 119–32. doi: 10.1080/0022250X.1971.9989791.
Sawyer, R.K. (2005), *Social Emergence: Societies as Complex Systems*. Cambridge: Cambridge University Press.
Schelling, T.C. (1971), "Dynamic models of segregation", *The Journal of Mathematical Sociology*, **1** (2), 143–86. doi: 10.1080/0022250X.1971.9989794.
Schelling, T.C. (1978), *Micromotives and Macrobehavior*, New York: W. W. Norton and Company.
Schneider, L. (ed.) (1967), *The Scottish Moralists: On Human Nature and Society*, Chicago, IL: University of Chicago Press.
Snijders, T.A.B., G.G. van de Bunt and C.E.G. Steglich (2010), "Introduction to stochastic actor-based models for network dynamics", *Social Networks*, **32** (1), 44–60. doi: 10.1016/j.socnet.2009.02.004.
Snijders, T.A.B. and R. Bosker (1999), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: SAGE.
Strogatz, S.H. and I. Stewart (1993), "Coupled oscillators and biological synchronization", *Scientific American*, **269** (6), 102–9.
Tajfel, H. and J.C. Turner (1986), "The social identity theory of intergroup behavior", in S. Worchel and W.G. Austin (eds), *Psychology of Intergroup Relations*, Chicago, IL: Nelson-Hall, pp. 7–24.
Turner, J.H. (1974), *The Structure of Sociological Theory*. Chicago, IL: The Dorsey Press.
Turner, J.H. (1995), *Macrodynamics. Toward a theory on the Organization of Human Populations*, New Brunswick, NJ: Rutgers University Press.
Udehn, L. (2002), "The changing face of methodological individualism", *Annual Review of Sociology*, **28**, 479–507. doi: 10.1146/Annurev.Soc.28.110601.140938.
Young, H.P. (1993), "The evolution of conventions", *Econometrica*, **61** (1), 57–84.
Young, H.P. (2015), "The Evolution of Social Norms", *Annual Review of Economics*, **7**, 359–87.

Michael Mäs - 9781789906851

5. Analytical sociology and social stratification research

Carlo Barone

1. INTRODUCTION

The Research Committee 28 (RC28) on social stratification and mobility of the International Sociological Association is a research community of sociologists analyzing social inequalities with quantitative methods. The RC28 is a success story within our discipline. It has evolved from a small group of social mobility researchers in the 1970s into the largest international network of social stratification scholars working on ethnic inequalities, gender segregation and socio-economic gaps in educational and occupational attainment, family dynamics and poverty, labor market exclusion and welfare inequalities, with a strong emphasis on longitudinal and comparative research. The RC28 is a dense network: it gathers leading scholars from European, North American and Asian universities who are actively involved in several joint projects and other forms of scientific collaboration. Crucially the RC28 is a rare instance of paradigmatic research in sociology (*sensu* Kuhn 1962) resulting in highly cumulative forms of knowledge: drawing on shared research questions, analytical frameworks and methodological tools, RC28 researchers have established several solid empirical generalizations concerning social inequalities in contemporary societies (Hout and Di Prete 2004; Goldthorpe 2006). Of course research on social inequalities extends well beyond the frontiers of the RC28 and, more generally, of quantitative sociology. However, in this chapter I will refer to this specific research community because, thanks to its visibility and internal coherence, it constitutes a well-recognized research tradition that can be meaningfully compared with analytical sociology (AS). In what follows, I refer to the former as Social Stratification Research (SSR), since its current institutional manifestations extend beyond the RC28.[1]

AS is a younger, smaller, more diverse community, covering a broader array of substantive domains and displaying more diverse theoretical and methodological orientations.[2] These two research networks have developed quite separately from each other and, as we will see, the limited intellectual exchanges between them have often involved some disputes about variable sociology and the relative importance of description and explanation in sociology. This is a paradoxical state of affairs. This chapter shows that the two research networks share a highly similar, and progressively convergent, research program. As discussed they share a commitment to solid descriptive *and* explanatory work, adherence to the principles of methodological individualism, a predilection for models of bounded rationality and a strong orientation toward deductive theory and hypothesis testing.

[1] For instance, there is a large degree of overlap between the RC28 and the ECSR (European Consortium for Sociological Research) in terms of membership to the two research networks and epistemological orientations, the main difference being that the latter includes several demographers studying social inequalities. Similar considerations can be made for other regional networks in Europe and North America stemming from the initial core of the RC28.

[2] It may be noted also that AS is a more informal network that is not institutionalized in any research committee within the International Sociological Association, and which does not have a reference scientific journal.

Carlo Barone - 9781789906851

While sharing a similar epistemological program for sociology, SSR and AS diverge with respect to the methodological and theoretical options to pursue it. I focus here on three issues: the preferred methods to establish the descriptive evidence concerning the *explananda*, the theoretical framework to develop explanatory models, and the empirical strategies to test these models. I will argue that significant differences in these three respects emerge not only between, but also *within* these two communities, and that some positions offer more promising options for the success of the broader research program shared by SSR and AS.

A significant hurdle for this discussion must be mentioned at the onset. It is difficult to make general statements about both AS and SSR, albeit for different reasons. As regards the former we can identify several programmatic articles and books proposing definitions of what AS is and is not, of what counts as a social mechanism, and of the best theories and methods to develop and test sociological explanations. While these programmatic statements display clear similarities, they also display some significant differences, and research practices within AS possibly reveal even more heterogeneity. For the purpose of characterizing the core elements of AS I will mainly refer to the programmatic statements found in the most representative, collaborative volumes published within AS (Hedström and Swedberg 1998; Hedström and Bearman 2009; Manzo 2014a).

We are confronted with a different kind of difficulty in the case of social stratification researchers who visibly display a limited appetence for programmatic statements. To circumvent this problem, I rely on Goldthorpe's (2016) volume on *Sociology as Population Science*, which is useful for this chapter because it proposes not only a description, but also a normative justification of the epistemological, theoretical and methodological orientations prevalent in SSR and in other domains of quantitative research in sociology, demography and epidemiology. Moreover, Goldthorpe is one of the most prominent scholars within SSR. Of course, his programmatic statements do not always correspond to the orientations and practices of RC28 researchers, as revealed in some respects by the following pages. But importantly, Goldthorpe has made core contributions to move SSR from a primarily descriptive exercise toward a theory-driven, explanatory research program. Along these lines he has developed an influential, formalized model of educational decisions grounded in the theory of bounded rationality (Breen and Goldthorpe 1997). Interestingly, this model displays significant affinities with explanations for educational inequality developed by analytical sociologists (Gambetta 1987; Manzo 2013) and their precursors (Boudon 1974). Hence, I will illustrate the main theses of this chapter using research on educational inequality, which is also my main field of expertise.

2. A SHARED RESEARCH PROGRAM FOR SOCIOLOGY

According to Goldthorpe (2016), sociology has a threefold mission: establishing solid, aggregate-level regularities of a probabilistic kind, identifying the individual-level mechanisms that generate them, and testing the explanatory power of the hypothesized mechanisms. In line with AS (Hedström and Swedberg 1998), and with explicit reference to some of its 'founding fathers' such as Merton, Coleman and Boudon, Goldthorpe stresses the importance of moving beyond statistical analysis to develop general narratives of action and interaction among individuals aimed at identifying the generative mechanisms producing the observed empirical regularities.

Carlo Barone - 9781789906851

At the same he stresses that sociological explanations should always start from well-established empirical *explananda*: it is just pointless to invest efforts at explaining phenomena that have not been well established. While this point may be easily recognized in principle, what needs to be stressed here is the importance of descriptive work with a highly *paradigmatic* orientation. While sociologists may agree that establishing solid descriptions is an important starting point to make good sociology, they often fail to realize that this objective demands a much more intensive and methodologically refined kind of descriptive empirical research than what is practiced in our discipline. This kind of research is very time-consuming and comes with a cost that most sociologists may not want to pay: narrowing the set of research questions that can be addressed against the temptation to claim that all social phenomena are interesting and important (Goldthorpe 2016).

Indeed, this kind of paradigmatic research is seldom practiced and is quite unpopular in sociology. SSR provides an interesting exception: a large community of scholars focusing on a relatively limited set of research questions, using highly standardized and progressively refined tools of data collection and statistical analysis, animated by the concern to establish a set of solid empirical regularities speaking to concerns for equality of opportunity and of condition in contemporary societies. This paradigmatic orientation is essential to establish solid explananda. For instance, considering educational inequalities, SSR has documented that (Erikson and Jonsson 1996; Hout and Di Prete 2004): (i) marked socio-economic gradients in educational attainment are consistently observed across different countries and historical periods; (ii) this result is highly robust to different measurements of family background and educational attainment, and to alternative modeling strategies; (iii) these gradients tend to decline across educational transitions, albeit in a non-monotonic way; (iv) these gradients are highly stable over cohorts: they moderately declined in the post-war decades in most western countries, but then stayed largely unchanged over the past four decades; (v) these gradients are smaller in countries displaying more comprehensive educational systems and more redistributive welfare states; (vi) these gradients in educational attainment are mediated by socio-economic gradients in academic achievement (primary effects), which start to develop in early childhood; (vii) family background affects educational enrollment decisions, even controlling for academic performance (secondary effects).

Without the intensive, prolonged and collaborative efforts of several researchers, it would have been impossible to identify this set of empirical regularities with a level of generalizability, robustness and detail that is seldom found in sociology. This kind of descriptive exercise is routinely discarded as misplaced empiricism in our discipline, including criticism against 'cameral sociology' (Boudon 2002) or 'variable sociology' (Hedström 2006). This can reflect a plausible concern that an excessive taste for methodological sophistication ends up replacing serious theoretical work. But what these critics often miss is that this collaborative, paradigmatic research strategy is necessary to arrive at the kind of well-established empirical regularities that constitute the ideal input of the explanatory models that AS aims at developing. Three considerations are particularly relevant in this respect.

First, a theoretical model should not aim at explaining a single empirical phenomenon, but rather several *interrelated* explananda. This is essential to avoid ad hoc explanations and to assess more rigorously the explanatory power of the theoretical model. The task of middle-range theories is precisely to federate multiple empirical findings into a coherent theoretical framework (Merton 1968). This should be the core empirical challenge determining whether an explanatory model succeeds or fails. Second, detailed descriptions are essential

Carlo Barone - 9781789906851

for explanatory purposes because they provide solid ground to *build* theoretical models. An opposition between descriptive and explanatory work is misplaced because the heuristic value of a theoretical model is highly contingent on the substantive plausibility of its assumptions and of the generative mechanisms that are selected for the explanation: making good choices demands fine-grained knowledge of the sets of constraints, opportunities and preferences of individuals. Third, once the explanatory model has been developed, its empirical testing also needs a good deal of empirical input, and once more detailed descriptions of interrelated phenomena are needed.

To illustrate, the theoretical model developed by Breen and Goldthorpe (1997) aims at explaining the existence of educational inequalities. However, after having identified three explanatory mechanisms, its authors mobilized them to account for the seven empirical regularities outlined, arguing that the ultimate success of their theoretical model should be determined on this basis. Similarly, the agent-based computational model of educational inequality proposed by Manzo (2009, 2013) provides a systematic explanation not only for the existence of educational inequalities in France, but also for their country and cohort variations across France and Italy. For this purpose, Manzo's model largely relies on the conceptual framework, methods, data and findings of educational stratification research. Moreover, this model rests on the assumption that families are sensitive to the costs, benefits and chances of success of educational investments—an assumption that had been repeatedly tested and validated in SSR (Breen and Goldthorpe 1997). Hence the selection of its generative mechanisms was largely guided by previous descriptive work. It may be noted that, in turn, this empirical work was inspired by theoretical models developed by Keller and Zavalloni (1964), Boudon (1974) and Gambetta (1987)—an excellent illustration of the potential complementarities between AS and SSR at the interface between description and explanation. Finally, Manzo (2013, p. 82) acknowledges as the main limitation of his model that 'the parameter values on which each micro-level mechanism relies only come from arbitrarily chosen numerical values'. Then Manzo recognizes that, for a more stringent test of the model, it would be useful to empirically calibrate it by inputting these parameters with survey data. Unfortunately, he lacked the relevant data for this purpose. More empirical data was needed for a more rigorous test of the theoretical model. Hence extensive descriptive analysis is needed to guide the selection of interrelated explananda, model building and model testing.

On one side, the importance of solid descriptive work has been increasingly recognized within AS (Manzo 2014a; Hedström and Bearman 2009). On the other side, SSR has largely abandoned past, simplistic claims about the supposed 'explanatory power' of path-analytic models (Boudon 1976) and, more generally, of statistics. Researchers have become increasingly aware that regression models cannot replace action-based, theoretical models. Within SSR mobility research has progressively evolved from an essentially descriptive exercise to a more theory-driven research field. The Breen-Goldthorpe model illustrates this tendency, but other examples can be mentioned within research on educational inequalities (Breen 1999; Breen and Yaish 2006; Erikson and Jonsson 1996; Compet and Berger 2017). At the same time Goldthorpe (2016) makes an important point when he urges stratification researchers to invest more efforts in explanatory work and theoretical advancement. As in most sociological work in this field, explanations often take the form of some vague, intuitive, ad hoc narrative to rationalize empirical findings.

I will return to this point. What I wish to stress here is that, overall, SSR and AS are progressively converging toward a shared understanding of the core tasks of our discipline:

Carlo Barone - 9781789906851

investing major efforts in establishing sets of detailed, interrelated empirical findings; moving beyond empirical data to identify their generative mechanisms via micro-level theories of individual behavior, following the principles of methodological individualism; developing rigorous empirical strategies to assess the heuristic power of the proposed explanations. At a deeper level this shared understanding of the mission of sociology probably reveals a common vision of sociology as a scientific exercise, which holds a meaningful degree of continuity and dialogue with other social and natural sciences—as opposed to postmodern, relativistic and subjectivist research programs orienting the discipline toward some sort of literary exercise (Goldthorpe 2016; Hedström and Bearman 2009). As discussed next, significant differences persist between AS and SSR, but they should not hide the fact that, within the highly balkanized or even decadent panorama of contemporary sociology, these two research communities have much in common. If their epistemological premises are highly similar, the *mise en oeuvre* of this shared research program reveals significant differences, starting with the priority attached to quantitative research and the standards to establish what counts as 'solid' evidence.

3. THE METHODOLOGICAL TOOLBOX FOR DESCRIBING SOCIAL INEQUALITIES: A PRIMACY OF SURVEY-BASED RESEARCH

According to Goldthorpe (2016, pp. 70, 77), sample survey research and multivariate data analysis have become central to sociology for the purpose of establishing solid descriptions. At the same time, grand historical sociology, post-modernist ethnography and qualitative case studies would have lost ground in our discipline. Goldthorpe argues that quantitative research is increasingly recognized as having 'premium payoffs' (2016, p. 1), thus attracting growing funding and an increasing number of sociologists.[3] This is rightly so according to Goldthorpe: for descriptive purposes, sociologists should accord a privileged position to survey data. Hence, statistical regularities should be the key *explananda* of our discipline. These regularities may be cross-sectional, but they may also refer to changes across the life course, cohorts or historical periods, as well as contextual effects relating to social networks, organizations or institutions (e.g. the influence of classmates, school characteristics and institutional arrangements of educational systems on educational attainment).

The prominence accorded to quantitative sociology certainly corresponds to the methodological orientations of SSR, but it is quite remote from the methodological pluralism characterizing AS (Hedström and Bearman 2009; Manzo 2014a). However, three observations are in order. First, Goldthorpe claims that survey data should be given priority *only* for descriptive purposes, while for explanatory purposes Goldthorpe encourages sociologists to mobilize a much larger toolbox: when it comes to testing explanatory mechanisms, experimental and quasi-experimental designs, agent-based models as well as theory-driven, qualitative case studies can be useful.

[3] Goldthorpe's perception of the success of quantitative research is maybe too influenced by what I earlier referred to as the success story of the RC28 and, perhaps, of few other domains of sociological inquiry. On the basis of his direct experience, Goldthorpe himself acknowledges that the UK is a relevant exception to this trend. On the same basis I would note that quantitative sociology remains persistently marginal in French and Italian sociology. And if the world conferences of the International Sociological Association give insights on the current methodological orientations of the discipline, a glance at conference programs suggests that an overriding majority of sociologists fails to see the 'premium pay-offs' of quantitative sociology.

Carlo Barone - 9781789906851

Second, Goldthorpe offers a compelling argument to argue that, for descriptive purposes, survey-based research entails important pay-offs: it provides a solid treatment to the problem of 'inferring from the part to the whole' (external validity). Importantly, this issue should not be seen as a concern of some specific research methods, but rather as a core challenge for any kind of empirical description. For well-known practical reasons, sociological research rarely involves the collection of data on the whole population of interest. Whatever method they use sociologists most often rely on samples. However, the results concerning a specific sample are not of any substantive interest if they cannot be used to make inferences to the larger population from which the specific sample is taken. The reason is simple: had we selected a different sample, we could have observed different results. Goldthorpe has, in my view, a strong point in stressing the fact that societies, social categories and social groups display high levels of individual-level *variability* with regard to most outcomes of interest to sociologists. The core advantage of statistical inferences based on probabilistic samples is that this individual-level variability is fully taken into account, resulting in *probabilistic* inferences which can avoid the systematic biases in the selection of cases that typically plague non-probabilistic sampling approaches prevalent in qualitative research (see Goldthorpe 2006: pp. 34–56 for a discussion of the pseudo-solutions to this problem offered in ethnographic research).

Against this argument one could reply that a qualitative study based on a small, non-probabilistic sample does not aim at providing any kind of macro-level generalization. This reply is not convincing, since then one could ask: what is the descriptive interest of the results that are thus obtained if results based on a different sample would have been different. Claiming that sociologists should be interested in the specific characteristics of a sample of 30 individuals is tantamount to condemning our discipline to irrelevance. A more convincing reply, according to Goldthorpe, is that qualitative studies can help explore the underlying processes driving pre-established, macro-regularities, provided that these studies are theoretically informed and motivated by explanatory purposes rather than sticking to some form of descriptive inductivism—a recommendation holding much affinity with the views of AS. It should be noted also that the target of Goldthorpe's criticism is not qualitative research per se. The same concerns about external validity can be addressed toward laboratory experiments, field experiments based on small, local and non-probabilistic samples and big data studies taken from highly selected samples. Whatever the method of inquiry, the problem of inferring from the sample to the population is always crucial, and survey research based on probabilistic sampling techniques offers the most convincing treatment of this problem.

My third remark on Goldthorpe's argument on the descriptive prominence of survey research is that it could (and should) be softened by noting that external validity is far from being ensured by probabilistic sampling. Goldthorpe acknowledges in a footnote (2016, p. 80) that declining, and increasingly biased, response rates can plague survey research, but he does not seem much concerned, because of the growing availability of methods of weighting and of multiple imputation of missing data. The problem is that these methods involve delicate assumptions that are difficult to test in most cases, such as the missing-at-random assumption (Rubin and Gill 1997). The biases arising from large, systematic non-response rates in survey research are attracting increasing attention and are an important reason for concern (Couper 2017). Hence, quantitative researchers may not want to oversell the virtues, nor hide problems, of survey research.

A more fundamental reason to soften Goldthorpe's claims on the supposed primacy of survey-based research is that external validity is only one of the parameters defining the

Carlo Barone - 9781789906851

quality of a study. The validity and reliability of measurements is an equally relevant concern, which can motivate sociologists to opt for other data collection methods. It is well known, for instance, that survey data can be problematic and unreliable for the purpose of measuring attitudes, beliefs and actions deviating from prescribed social norms. For instance, the study of decision-making processes in organizations based on survey questionnaires can easily result in highly idealized, socially desirable representations which do not correspond much to the real functioning of organizations. In this case, triangulating direct observations, qualitative interviews and document analysis can be a more fruitful descriptive exercise. However, since this kind of research strategy cannot be easily implemented on a large, probabilistic sample of organizations, a researcher could reasonably decide to trade external validity for more ecologically valid, reliable observations. Even in this context—I would stress—the problem of external validity should be taken seriously. Researchers should carefully describe and justify the selection of cases, spell out its potential limitations for the generalizability of findings, and then replicate the same study across different samples. Repeated observations based on shared protocols across different contexts would be equally relevant in a qualitative context, which in turn involves moving toward more collaborative, coordinated, intensive research practices—qualitative research could and should be as paradigmatic as quantitative research.

Issues of internal validity are no less important. They provide further reasons to enlarge the toolbox of methods for the social sciences. I can illustrate this point with studies on employer discrimination. We are confronted here with another example of measurement issues plaguing survey research: administering a questionnaire to either employers or applicants to assess discriminatory practices is clearly problematic, as we are most likely to obtain information only on perceived discrimination. Indeed, in survey research the most common strategy is to indirectly infer discrimination by collecting data on applicants' characteristics and skills that are relevant for the job position for which they apply. If in controlling for these characteristics we observe that the disadvantaged category (women, ethnic minorities, working-class individuals) still holds lower chances of being hired, we may want to infer that this category is discriminated (Neumark 2012). However, this inference is highly problematic since it relies on the heroic assumption that *all* characteristics relevant to the performance have been incorporated in the statistical model. Correspondence studies provide a more solid treatment to this classical problem of omitted-variable bias (*ibidem*). In standard applications matched pairs of qualitatively identical job applications are sent to employers who have advertised a job opening; the only difference between these applications refers to the names or the photographs of the fictious applicants, which indirectly signal ethnicity, gender or social origins. Discrimination in hiring is then quantified by estimating group differences in call back rates (i.e. invitations to a job interview). Crucially, by *randomly* assigning equivalent applications to employers we can avoid the risk that results are biased by systematic differences in observed or unobserved characteristics of the applicants.

Since omitted-variable bias is a very general problem, issues of internal validity are pervasive in every domain of empirical research. Another example in stratification research involves mobility studies assessing the association between education and occupational attainment, its cohort and country variations. There is a long list of genetic, personality, contextual and socio-economic factors affecting educational attainment, which can also affect occupational attainment. If some of these confounders are omitted, the resulting estimates of 'returns to education' can be severely biased. I am not aware of any a single correlational study based on survey data controlling for more than just a few of these confounders. The reason is simple:

Carlo Barone - 9781789906851

incorporating all the potential confounders supposes that all determinants of educational attainment affecting occupational attainment are *known* to the researcher, and that they have been *measured* in survey data. Both prerequisites are highly unrealistic.

Goldthorpe (2016) invokes a common defensive argument to deal with this problem: correlational studies based on survey data would only aim at assessing associations, rather than establishing causality. This argument is not very convincing because, at the same time, Goldthorpe argues that multivariate statistics are important precisely for the purpose of controlling for potential confounders. He thus explicitly acknowledges that sociologists should neutralize spurious correlations arising from unobserved confounders, when describing empirical regularities. The problem is that, unfortunately, most sociologists fail to recognize that the list of potential confounders is always much longer than the list of control variables available in survey data. I wish to stress again that qualitative research is not exempt from these issues: ethnographic results on micro-level processes accounting for any given social phenomenon can be misleading if the researcher misses other parallel processes affecting the same phenomenon correlating with the observed processes. Ethnography does not open any immediate window into causality.

Field and laboratory experiments, as well as quasi-experimental techniques, such as regression discontinuity and difference-in-difference designs, provide a more convincing treatment of the issue of internal validity. Randomized designs neutralize all potential confounders, either known or unknown to researchers, and measured or unmeasured in their data (Morgan and Winship 2007). At the same time these alternatives to survey-based statistical research should not be seen as a 'golden standard' that must replace it. They can present significant limitations with respect to their external validity or the quality of their measurements (for instance, the failure to assess long-term impacts in several experimental studies). As for qualitative research, researchers can reasonably decide that these 'costs' are worth paying and, again, they should anyway direct all their efforts to minimize them. For instance, in my own experimental studies I have criticized previous field experiments relying on small ad hoc local samples, and I have used multi-site experimental designs based on probabilistic samples (Barone et al. 2018, 2019b).

More generally, different research designs entail different trade-offs between concerns for external validity, internal validity and measurement. The important consideration is that the relevance of *all three criteria* should be recognized *whatever method of data collection* and analysis is chosen. Accordingly, researchers should direct their efforts to obtain the best compromises between these criteria. This is, in my view, the dividing line between good and bad empirical research.

Hence, the recourse to quantitative methods in SSR should not be accorded any normative status. Methods are more or less appropriate depending on the specific research questions at hand. Certainly the large recourse to survey data in SSR reflects a consensus on the importance of external validity and the virtues of probabilistic sampling. However, this orientation also reflects the specific kind of research questions addressed in this field. In particular, social mobility research has extensively engaged in the task of testing modernization theories predicting variations in educational and occupational inequalities driven by industrialization and tertiarization: the assessment of these hypotheses involved large-scale comparisons between representative samples of different countries and cohorts (Goldthorpe 2006). Once these comparisons revealed significant variations that could not be explained by modernization theory, SSR started to explore the role of institutional arrangements in educational systems (e.g. tracking

Carlo Barone - 9781789906851

age, standardization) and in welfare states for these patterns of social inequalities (Shavit and Müller 1998). Again, country and longitudinal comparisons were of critical importance, and they supposed reliable inferences about the cohort- and country-level patterns of inequality to be compared. The high external validity of survey research is a major consideration given these research questions, and the measurement of the few core variables involved in these analyses (origins, education, occupational destinations) is regarded as reasonably valid and reliable. At the same time, the growing awareness of the problems of internal validity within SSR has motivated a rapid increase in the recourse to experimental and quasi-experimental designs (Barone and Solga 2020),[4] which echoes the frequent use of experimental methods in AS (Hedström and Bearman 2009).

4. THE THEORETICAL TOOLBOX FOR EXPLAINING SOCIAL INEQUALITIES: BEYOND RATIONAL CHOICE THEORY

We have seen that AS advocates more systematic explanatory efforts to identify the mechanisms underlying social phenomena and that over the past two decades SSR has increasingly moved in this direction. Moreover, researchers from both sides, adhering to the prescriptions of methodological individualism, claim that an adequate understanding of these mechanisms demands individual-level, micro-founded explanations (Hedström and Swedberg, 1998; Hedström 2006). Even more specifically, in both research communities several scholars have often mobilized rational choice theory (RCT) for explanatory purposes (Hedström and Swedberg 1996; Breen and Goldthorpe 1997; Erikson and Jonsson 1996).

Within the broader framework of methodological individualism, RCT introduces the additional postulate of consequentialism, thus assuming that human beings are able to make informed choices on the basis of their capacity to prefigure alternative courses of action and to assess their potential consequences, given the ends that they pursue (Hechter 1996; Boudon 1999; Goldthorpe 2016).

Within consequentialist models, the expected utility theory introduces four additional, restrictive assumptions: the relevant ends involve the utility for the individual of the alternative courses of action (self-interested motives); this utility is most often identified with reference to economic returns (e.g. lifetime earnings); decision-makers have complete and unbiased information on these consequences (perfect information); and they effectively process this massive amount of information, thus selecting the course of action maximizing individual utility (maximization).

Virtually all sociologists, within and outside SSR and AS, reject these restrictive assumptions as unrealistic. Several sociologists, then, tend to reject RCT *in toto*, mostly because they identify it with economic RCT, and they worry about economic imperialism in the social

[4] Goldthorpe correctly notes that these methods should not be seen as substitutes for theory-driven, individual-level explanations. I have employed the term 'internal validity' rather than 'causal inference' precisely to avoid any misunderstanding about 'causality'. My point is that counterfactual approaches can provide a sound treatment of classical problems of omitted-variable bias and reverse causality plaguing standard multivariate statistics. Unfortunately, Goldthorpe discards this approach as a form of applied research focusing on the effects of policy interventions, thus revealing a narrow understanding of these methods, as suggested by the example on correspondence studies to measure employer's discrimination and, more generally, by a broad range of experimental studies across several fields of stratification research (Barone and Solga 2020).

Carlo Barone - 9781789906851

sciences (see Coleman and Fararo 1992). They also often criticize RCT as an unduly atomist conception of individual action, which is again an arguably sound critique of several economic applications of RCT (*ibidem*).

However, other sociologists reject the claim that sociologists should regard rational action as a prerogative of economists. Indeed, following in the footsteps of classical sociologists such as Marx, Pareto and Weber, they argue that instrumental rationality is highly pervasive, not just in the realm of economic life (Hechter and Kanazawa 1997). Crucially, they note that consequentialism is a far more general conception of rationality than the narrow version of RCT adopted in mainstream economics (*ibidem*). Self-interested individuals may attach importance to rewards other than economic or even material benefits, such as power and prestige. More importantly, sociological versions of RCT recognize that individuals are not always self-interested. It is simply impossible to explain all forms of altruistic or normative behavior as stemming from some hidden, self-interested motives. The important point, then, is that consequentialism is a far more general conception of human action than self-interested behavior (Hechter 1996; Goldthorpe 2006). For instance, even when we are genuinely motivated by the concern to help other individuals we are often sensitive to the costs that we may incur and/or to the benefits that others will reap: we are pragmatic altruists. Similarly, RCT is compatible with the possibility that individuals follow some cultural routines or social norms, to the extent that, in doing so, they do not act as 'cultural idiots' blindly following cultural scripts, but rather retain some degree of autonomy from sociocultural conditionings (Hechter and Kanazawa 1997; Goldthorpe 2016, pp. 54–66). Furthermore, most sociological versions of RCT reject the assumptions of perfect information and infinite computational ability, in favor of more realistic models of bounded rationality, arguing that individuals often hold inaccurate or even biased beliefs about the consequences of their actions (*ibidem*). In these sociological applications detailed cost-benefit calculations are thus replaced by the 'fast and frugal heuristics' observed by psychologists when analysing decision-making processes

What is perhaps most characteristic of *sociological* applications of RCT is the rejection of the atomist conception of action characterizing several economic applications of RCT (Hedström and Bearman 2009; Goldthorpe 2016). In sociological analyses of decision-making processes, the choice set, as well as the related configurations of opportunities and constraints, can systematically vary among individuals, depending on their embeddedness in different institutional environments, social classes, status groups and social networks. Most sociologists mobilizing RCT thus adhere to the 'structural' version of methodological individualism that focuses on the interplay between social structures and individual-level decisions. This is indeed the variant of methodological individualism most often practiced within SSR and AS (Goldthorpe 2016; Manzo 2013).

These theoretical moves enhance considerably the realism of RCT, its range of applicability and its sociological relevance. Instrumental rationality is too important to be ignored by sociologists: RCT provides a coherent, clear and parsimonious framework to incorporate it in sociological explanations. At the same time these moves raise important difficulties, and SSR and AS have responded differently to these difficulties. While in the former RCT is still accorded a privileged status for the reasons that I just mentioned, several adherents to AS have increasingly stressed the limitations of consequentialist explanations, including sociological variants of RCT. Let us see why.

For many sociologists, an important justification for adopting RCT is that rationality is an intuitively plausible motive of action and thus needs not to be explained: as effectively

Carlo Barone - 9781789906851

summarized by Hollis (1997), RCT is its own explanation. This is arguably because we intuitively realize that the capability to explore the potential consequences of our actions helps us make better decisions and adapt to our environment. Therefore, it is unsurprising that we are capable of acting rationally. More explicitly the existence of rationality can be explained in evolutionary terms: developing a capability to realize that alternative courses of action are possible, to anticipate some of their consequences, and to understand that some options are more advantageous than others ensured huge pay-offs for the survival of our ancestors (Haidt 2012). In this perspective, it also becomes understandable that our capability to anticipate the consequences of our actions is highly *rudimental*, since in the ancestral environment where our brain evolved we did not need very detailed information nor highly sophisticated computing skills, but rather a pragmatic ability to anticipate the most important, proximate opportunities and dangers, and to rapidly identify some feasible, adaptive response (Kahnemann 2011).

However, it is much less self-evident why and how altruist behavior (other than kin altruism) and loyalty to group norms developed (Haidt 2012; Gintis et al. 2003). While they were probably adaptive for the survival of the group, they were not adaptive from an individual perspective:[5] free-riding is a more adaptive strategy. Hence, the argument that RCT is its own explanation becomes problematic once we enlarge the set of preferences compatible with consequentialism to make the theory more realistic. While the existence of consequentialism is self-evident from an evolutionary perspective, altruism and group loyalty represent an unsolved mystery from the same perspective.

The issue is not simply about the origins of rationality and self-interest. The most compelling problem for empirical research is that, if the mechanisms driving 'deviations' from self-interest remain a black box, RCT cannot explain why in some situations individuals behave altruistically and in others they do not, or why and under which conditions they adhere to group norms. Variations in preferences are left unexplained in RCT: *de gustibus non est disputandum*. Similarly, RCT cannot provide any systematic explanation for variations in the beliefs about the consequences of our actions, other than tracing them to the utilitarian pursuit of some hidden interest. While this can be an important consideration, it falls short of developing an adequate explanation for variations in human beliefs. In sum, once the existence of systematic variations in individual preferences and beliefs is acknowledged, the problem is how to explain these variations. The lack of an adequate theory of preferences and beliefs within RCT to account for these variations undermines the explanatory power and the credibility of the theory.

In SSR these difficulties are either ignored or are regarded as ineluctable limitations of any theory of action. This is the position taken by Goldthorpe (2016): while recognizing that individual preferences can systematically vary across time and space, he recommends sociologists following economists treat these preferences as (unexplained) 'givens'. Individual ends should thus be treated as exogenous. Goldthorpe's position also illustrates the two most common defensive arguments mobilised to justify this position. In order to accommodate variations in individual preferences (or beliefs) one can argue that they are randomly distributed among individuals; maybe this assumption is too strong, but possibly not too far from

[5] Group-level natural selection is taboo in evolutionary theory for a number of empirical and theoretical reasons that have been extensively reviewed elsewhere (Haidt 2012). However, the debate is still open, and some argue that it is not necessarily a wrong hypothesis.

Carlo Barone - 9781789906851

reality. Then, these variations may be discarded as largely irrelevant to sociological explanations. The other defensive move is to rationalize altruist or normative behavior as hidden forms of consequentialism. For instance, individuals may follow social conventions to spare the costs of experimenting and learning, or they may adhere to social norms to overcome coordination problems (Goldthorpe 2016, p. 221). Both arguments boil down to the same conclusion: variations in preferences and beliefs could and should be removed from sociological analysis.

The Breen-Goldthorpe (1997, p. 278) model exemplifies this position, since its authors 'seek to dispense with any assumption that actors will be subject to systematic influences of a (sub)cultural kind, whether operating through class differences in values, norms or beliefs regarding education or through more obscure "subintentional" processes', thus pursuing 'an undoubtedly reductionist' (*ibidem*, p. 300) approach to sociological explanation. However, further theoretical elaborations of this model indicate that there are valid reasons to expect systematic, class-based differences in beliefs about education (Erikson and Jonsson 1996; Breen 1999; Breen and Yaish 2006). Moreover, a growing number of field experiments provide strong evidence that these class-based differences in beliefs can fuel educational inequalities (Barone et al. 2018, 2019; Herbaut and Geven 2020), thus challenging the reductionist ambitions of the Breen-Goldthorpe model.

More fundamentally, deviations from consequentialism—the core postulate of RCT—are too substantively important to be dismissed. The paradox of voting illustrates this problem (Boudon 1999). As the probability to exercise the pivotal vote in an otherwise tied election is virtually zero, consequentialist explanations of voting are problematic. Even bounded rationality models, invoking some cognitive illusion concerning the perceived costs, probabilities or benefits of exercising the pivotal vote, empirically fail to fully explain why people vote (Hill 2004). We may of course explain this anomaly by postulating some intrinsic utility of voting, but this kind of ad hoc explanation is little more than rephrasing the problem. Instead, sociologists may want to acknowledge that individuals *sometimes* have subjectively meaningful reasons for their actions that are unrelated to any plausible consequence (Boudon 1999).

Against this background some analytical sociologists consider that it is more realistic to abandon the claim that RCT, even in a bounded rationality formulation incorporating altruistic preferences, provides a general theory of action (Hedström and Bearman 2009; Manzo 2014a). Then, rather than working at a more comprehensive theory of action, some of them prefer 'moving down' in terms of abstraction and generality to identify a set of specific individual-level mechanisms providing the 'nuts and bolts' of sociological explanations (Hedström and Swedberg 1998; Hedström and Bearman 2009). These include, for instance, self-fulfilling prophecies, Matthew effects, social influence, or cognitive dissonance. Hence, middle-range theories are regarded as the most feasible, realistic option for sociology. However, as noted by Goldthorpe (2016, p. 22), the theoretical integration that is thus pursued is far from satisfactory: sociologists are left with a long *catalogue raisonné* of disconnected mechanisms. The concrete risk is that they end up picking ad hoc the explanatory mechanisms with limited substantive justification.

Overall, the behavioral assumptions of RCT are too narrow for a general theory of action; its failures to endogenize preferences and beliefs are too problematic to be ignored. At the same time, the less ambitious, but more flexible, mechanism approach comes with the price of theoretical fragmentation, and with the risk of formulating ad hoc explanations: the conditions of activation of specific mechanisms remain under-specified.

It is clearly outside the ambitions of this work to spell out a solution to these problems. I will only add four comments. First, AS proposes a sociologically relevant solution to analyze

Carlo Barone - 9781789906851

variations in preferences and beliefs by considering the influence of social networks. For actors with bounded rationality it is plausible to regard their *social milieu* as a source of information and as a benchmark for their aspirations. For instance, Manzo's model (2013) integrates standard RCT models with an analysis of peer influences on educational decisions. Second, if the above remarks are correct, they suggest that a promising avenue is to work at a general theory of action based on broader behavioral assumptions than RCT. Third, this theory is unlikely to emerge if empirical sociologists fail to engage in more systematic explanatory efforts, making the behavioral assumptions of their models more explicit. My expectation is that a more general theory can only emerge from the convergence between empirical research and theoretical debates. Fourth, a promising avenue is to integrate models of instrumental rationality with more general theoretical models specifying the conditions under which individuals may be animated or not by self-interested motives and they may entertain or not accurate beliefs. Dual-process theories are one such promising candidate (Kahneman 2011; Kroneberg et al. 2010, Kalter and Kroneberg 2014). Instead of taking rational action as the rule, they suggest that individual behavior oscillates between two analytical poles. Our actions are routinely guided by System 1 processes, characterized by simple heuristics guided by automatic processes providing fast, low-effort, pragmatic solutions to most of our decision-making problems. Importantly, this theory then specifies the conditions under which more reflexive, analytical System 2 processes emerge, including consequentialist behavior guided by the assessment of costs and benefits. For instance, rational behavior is more likely to occur when decision processes are diffused over time and involve repeated choices so that we can learn from experience; the stakes are perceived as high and the information needed to arrive at a decision is available at a relatively low cost. Since educational decisions tend to meet these criteria, RCT may provide a reasonable approximation of educational decision-making. At the same time dual-system models can be mobilized to explain under which conditions biased beliefs can affect inequalities in education and how these belief-driven mechanisms can be contrasted by counseling intervention (Barone et al. 2019b).

5. TESTING EXPLANATIONS OF SOCIAL INEQUALITIES: THE IMPORTANCE OF FORMALIZATION

The practice of *formalizing* sociological explanations is more common in AS and SSR than in most other areas of sociological inquiry, but it remains a rather infrequent practice even in these two research communities, perhaps more in the latter. Unfortunately, most sociologists fail to appreciate the virtues of formalizing theoretical explanations. Transparency is a fundamental value in scientific work; verbal narratives incur all kinds of ambiguities and fallacies to meet the standards of rigorous theoretical work. Transparency is essential because it is only if we clearly spell out the assumptions behind sociological explanations and how they lead to specific substantive predictions that these assumptions can be openly scrutinized, criticized and replaced with better assumptions.

As an illustration, in my own work I have contrasted the mathematical formalization of the BG model with its verbal formulations given by the authors and found in the most-cited articles using this model (Barone et al. 2019b). The mathematical model assumes that when assessing the occupational benefits of educational investments families are *exclusively* driven by the concern to avoid downward mobility. Hence, families would not differentiate between immobility and long-range upward mobility; these options would entail the same utility, which is highly

Carlo Barone - 9781789906851

unrealistic (and maybe even irrational). However, the systematic review of empirical tests of the BG model indicates that this assumption is systematically misreported in verbal accounts of the model, where it is replaced by the more generic, plausible claim that families *prioritise* avoiding downward mobility. This mismatch has important consequences. On one side, using mathematical formalization we show that this restrictive assumption leads to implausible predictions: for instance, according to this model families of unskilled manual workers (the 'underclass') hold the highest level of educational aspirations, which is implausible. On the other side, comparing the BG model with the mathematical reformulation proposed by Breen and Yaish (2006), we show that, when the restrictive assumption is replaced with more realistic assumptions, the Breen–Goldthorpe model leads to different, weaker substantive predictions (Barone et al. 2019b).

This example illustrates three points. First, when model assumptions are formalized, they can be closely scrutinized and criticised because they are unambiguous. Verbal accounts are more opaque to criticism. Second, empirical researchers often attach limited importance to formalization, failing to appreciate how slight differences in the verbal formulation of model assumptions can have important implications for model predictions. My third remark is that the benefits of transparency ensured by mathematical formalization often come with a cost: mathematical tractability often demands simplifying unrealistic assumptions. This is a most likely reason for the introduction of the mentioned unrealistic assumption of the BG model. The Breen–Yaish (2006) reformulation shows that the degree of realism can be progressively enhanced, but only within certain limits. For instance, modelling interactions between individuals with mathematical models, as done in game theory, can demand highly restrictive, dubious assumptions (Macy and Flache 2009). Agent-based computational models provide a more flexible alternative to mathematical models to formalize theoretical explanations, particularly when modelling interactions. Unsurprisingly, these models are particularly popular in AS, which tends to stress the importance of interaction mechanisms and the explanatory relevance of social networks. These advantages come with important costs, too. Unlike mathematical models, the results of agent-based models are highly dependent on the specific numerical values that are imputed. As noted by Macy and Flache (2009, p. 262), 'the number of numerical combinations explodes as the number of parameters increases, making a full exploration of the parameter space impractical'.[6] As mentioned above, empirical calibration can be a partial remedy, but researchers often lack the relevant survey data to implement it.

Even when this data is available, empirical calibration faces the problems of internal validity of survey data discussed in section 3: the results of the simulation can be biased by the large number of omitted confounders affecting the empirical caliper. Calibration based on experimental or quasi-experimental evidence would be a promising avenue (Boero and Squazzoni 2005), but then issues of external validity and data availability become even more pressing. Overall, different approaches to model formalization entail different costs and benefits. These options, together with their costs and benefits, should be systematically considered by empirical researchers interested in doing serious theoretical work.

[6] If the number of parameters is relatively small, some techniques to explore the parameter space can attenuate this problem (for a discussion of this important point, see Manzo (2014b)).

Carlo Barone - 9781789906851

6. CONCLUDING REMARKS

This chapter has argued that SSR and AS share similar research programs, which could further converge around a threefold mission for sociology: (i) developing intensive, collaborative efforts to establish sets of solid, interrelated empirical findings, which provide the explananda for theoretical models, guide model selection and model testing; (ii) developing theoretical models based on versions of structural individualism that integrate instrumental rationality into a broader theoretical framework incorporating non-consequentialist logics of action and stressing the explanatory role of networks and institutions to account for systematic variations in individual preferences and beliefs; (iii) testing the explanatory power of these models by deriving unambiguous predictions based on the formalization of generative mechanisms, rather than only on verbal narratives. If this proposal is feasible and promising, there are good reasons to acknowledge the complementarity of SSR and AS and to develop more systematic connections and collaborations.

Moreover, while existing differences between AS and SSR should not be hidden, they need not be elevated to normative status. For instance, I have argued that, rather than advocating the supposed primacy of specific methods of inquiry to describe social phenomena, sociologists could acknowledge that some core methodological criteria (internal validity, external validity, validity and reliability of measurements) are important whatever method is used; they would thus select the research designs providing the best compromises among them depending on the specific research questions. Similarly, rather than stubbornly sticking to RCT despite its above-outlined limitations, or accepting the theoretical fragmentation resulting from the 'catalogue approach' to mechanism-based explanations, sociologists could work at developing more comprehensive theoretical frameworks, which must specify the contextual conditions under which some logics of action are more likely to emerge. If this ambitious objective demands a growing integration between theoretical and empirical work, it provides further reasons to advocate more systematic collaboration between SSR and AS.

REFERENCES

Barone C., G. Assirelli, G. Abbiati and G. Argentin (2018), "Social origins, relative risk aversion and track choice", *Acta Sociologica*, **61** (4), 441–59.

Barone, C., K. Barg and M. Ichou (2019a), "Relative risk-aversion models: How plausible are their assumptions", OSC working paper, 3/2019.

Barone, C., A. Schizzerotto, G. Assirelli and G. Giovanni (2019b), "Nudging gender desegregation: A field experiment on the causal effect of information barriers on gender inequalities in higher education", *European Societies*, **21** (3), 356–77.

Barone, C. and H. Solga (2020), "Experimental methods in social stratification research", *Research in Social Stratification and Mobility*, **36** (2).

Boero, R. and F. Squazzoni (2005), "Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science, *Journal of Artificial Societies and Social Simulation*, **8** (4).

Boudon, R. (1974), *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society*, New York: Wiley.

Boudon, R. (1976), "Comment on Hauser's review of education, opportunity and social inequality", *The American Journal of Sociology*, **8** (5), 1175–87.

Boudon, R. (1999), *Le sens des valeurs*, Paris: Presses Universitaires de France.

Boudon, R. (2002), "Sociology that really matters", *European Sociological Review*, **18** (3), 371–8.

Breen, R. (1999), 'Beliefs, rational choice and Bayesian learning', *Rationality and Society*, **11** (4), 463–80.

Breen, R. and J.H. Goldthorpe (1997), "Explaining educational differentials towards a formal rational action theory", *Rationality and Society*, **9** (3), 275–305.

Carlo Barone - 9781789906851

Breen, R. and M. Yaish (2006), "Testing the Breen-Goldthorpe model of educational decision making", in S.L. Morgan, D.B. Grusky, G.S. Fields (eds), *Mobility and Inequality: Frontiers of Research in Economics and Sociology*, Stanford, CA: Stanford University Press, pp. 232–58.
Coleman, J. and T. Fararo (1992), *Rational Choice Theory: Advocacy and Critique*, Newbury Park: SAGE.
Combet, B. and J. Berger (2017), "Late school tracking, less class bias in educational decision-making? The uncertainty reduction mechanism and its experimental test", *European Sociological Review*, **33** (1), 124–36.
Couper, M. (2017), "New developments in survey data collection", *Annual Review of Sociology*, **43**, 121–45.
Erikson, R. and J. Jonsson (1996), *Can Education Be Equalised? The Swedish Case in Comparative Perspective*, Boulder, CO: Westview Press.
Gambetta, D. (1987), *Were they Pushed or did they Jump?* New York: Cambridge University Press.
Gintis, H., S. Bowles, R. Boyd and E. Fehr (2003), "Explaining altruistic behavior in humans", *Evolution and Human Behavior*, **24** (3), 153–72.
Goldthorpe, J.H. (2006), *On Sociolo*gy, Oxford: Oxford University Press.
Goldthorpe, J.H. (2016), *Sociology as a Population Science*, Cambridge: Cambridge University Press.
Haidt, J. (2012), *The Righteous Mind*, New York: Pantheon Books.
Hechter, M. and S. Kanazawa (1997), "Sociological rational choice theory", *Annual Review of Sociology*, **23** (1), 191–214.
Hedström, P. (2006), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge: Cambridge University Press.
Hedström, P. and P. Bearman (2009), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press.
Hedström, P. and R. Swedberg (1996), "Rational choice, empirical research, and the sociological tradition", *European Sociological Review*, **12** (2), 127–46.
Hedström, P. and R. Swedberg (1998), *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge: Cambridge University Press.
Herbaut, E. and K. Geven (2020), "What works to reduce inequalities in higher education? A systematic review of the (quasi-)experimental literature on outreach and financial aid", *Research in Social Stratification and Mobility*, **65**, 100442, https://doi.org/10.1016/j.rssm.2019.100442.
Hill, L. (2004), "Compulsory voting in Australia: A basis for a 'best practice' regime", *Federal Law Review*, **32** (3), 479–97, https://doi.org/10.1177/0067205X0403200307.
Hollis, M. (1977), *Models of Man*, Cambridge: Cambridge University Press.
Hout, M. and T. Di Prete (2004), "What we have learned: RC28's contributions to knowledge about social stratification", http://www.columbia.edu/~tad61/Hout&DiPrete_23July04.pdf.
Kahneman, D. (2011), *Thinking, Fast and Slow*, New York: Farrar.
Kalter, F. and C. Kroneberg (2014), "Between mechanism talk and mechanism cult: new emphases in explanatory sociology and empirical research", *Koelner Z. Soziologie*, **66**, 91–115.
Keller, S. and M. Zavalloni (1964), "Ambition and social class: A respecification", *Social Forces*, **43** (1), 58–70
Kroneberg, C., M. Yaish and V. Stocké (2010), "Norms and rationality in electoral participation and in the rescue of Jews in WWII: An application of the model of frame selection", *Rationality and Society*, **22**, 3–36.
Kuhn, T. (1962), *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.
Macy, M. and A. Flache (2009), "Social dynamics from the bottom up: Agent-based models of social interaction", in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press.
Merton, R. (1968), *Social Theory and Social Structure*, New York: Free Press.
Manzo, G. (2009), *La spirale des inégalités: choix scolaires en France et en Italie au XXe siècle*, Paris: Presses de l'université Sorbonne.
Manzo, G. (2013), "Educational choices and social interactions: A formal model and a computational test", *Comparative Social Research*, **30** (Class and Stratification Analysis), 47–100.
Manzo, G. (2014a), "Data, generative models, and mechanisms: More on the principles of analytical sociology", in G. Manzo (ed.), *Analytical Sociology: Actions and Networks*, Chichester, UK: Wiley and Sons, pp. 1–52.
Manzo, G. (2014b), "The potential and limitations of agent-based simulations", *Revue Française de Sociologie*, **55** (4), 653–88.
Morgan, S. and C. Winship (2007), *Counterfactuals and Causal Inference*, Cambridge, MA: Harvard University Press.
Neumark, D. (2012), "Detecting discrimination in audit and correspondence studies", NBER Working Paper No. 16448.
Rubin, J. and R.D. Gill (1997), "Non-response models for the analysis of non-monotone ignorable missing data", *Stat. Med*., **16**, 39–56.
Shavit, Y. and W. Müller (eds) (1998), *From School to Work*, Oxford, UK: Oxford University Press.

Carlo Barone - 9781789906851

6. Analytical sociology and critical realism

Francesco Di Iorio and Francisco J. León-Medina

1. INTRODUCTION

Critical realism, a philosophical orientation that during the last few decades has influenced eminent epistemologists, sociologists and economists, provides a theory of the social sciences that shares many features with analytical sociology. In particular, these two approaches agree on the idea that social systems are open and indeterminist; on the inadequacy of both positivism and relativism; on the fact that social phenomena are based on generative mechanisms; and also on the limits of both holism and atomism. The fact that they hold so much in common might have tempted us to suppose that critical realists would be sympathetic to the aims and methods of analytical sociology. Instead, they have been forthright in their criticisms. In this context, a systematic analysis of the relationship between these two approaches, as well as of the criticisms leveled by critical realists at analytical sociology, is overdue. In the following pages we will attempt to make good this deficit. Our intention is to defend the research program of analytical sociology from the attacks of critical realists. Moreover, we will discuss some of the shortcomings of critical realism as a theory of science.

The task of comparing analytical sociology with critical realism is challenging not only because of the scarcity of existing accounts of their connections, but also because of the different intellectual, philosophical and lexical backgrounds of the two approaches. We will proceed in the following stages. Section 2 presents the ontological assumptions of critical realism that are strictly related to a foundationalist theory about the unobservable deep structures of reality. Section 3 shows that this foundationalist theory is philosophically problematic and highlights its incompatibility with analytical sociology. The criticisms that critical realists make against analytical sociology because of its commitment to methodological individualism (MI) will be the focus of section 4. Section 5 debunks these criticisms, which equate MI with atomism, clarifying that analytical sociologists use a non-atomistic variant of MI that allows for a structural analysis of social phenomena. Section 6 tackles the differences between the analytical and the critical realist conceptions of social mechanisms and causation. Finally, section 7 presents the main differences between the research agendas of analytical sociology and critical realism.

2. CRITICAL REALISM AS AN ONTOLOGY OF THE DEEP STRUCTURES OF REALITY

Critical realism's philosophical approach to the nature of science and its assumptions was first developed in the UK in the 1970s, under the influence of Roy Bhaskar (1944–2014). Bhaskar and his followers, such as Margaret Archer, Philip Gorski and Tony Lawson, consider critical realism to be a reaction to the shortcomings of positivism, on one hand, and those of

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

interpretativism on the other (Gorski 2013, p. 659). Critical realists defend an ontology of science that is rooted in Marx's theory of ideology. Like the latter theory, critical realism reduces the relevance of empirically observable data in science and stresses the necessity of going beyond appearances to grasp the deep structures of reality that escape our senses (Bhaskar and Callinicos 2015; Turner 2013, p. 417).

The limitations of positivism, according to critical realists, derive from its being a form of "naïve or commonsense realism" (Gorski 2013, p. 659). As is well known, positivism assumes that scientific knowledge takes the form of testable "general or covering laws", i.e. "universal and exception-less statements that enable us to predict and control events" (p. 661). According to critical realists, many phenomena cannot be accounted for in terms of constant and necessary conjunctions because explanations in terms of exception-less laws only apply "under the special conditions of a controlled experiment" (Chalmers 1988, p. 18), i.e. under artificial laboratory conditions. In their opinion, real world phenomena are not governed by simplistic and deterministic patterns: "Autumn leaves rarely, if ever, fall to the ground in accordance with the law of fall" (p. 18). Critical realists advocate transcending the positivist ontology implicit in the constant conjunctions account and explaining phenomena instead in terms of "generative mechanisms" that lie behind the observable world (see Bhaskar (1975) [1997]). In their view, generative mechanisms are "trans-empirical", but nevertheless real existing entities "explaining why observable events occur" (Blom and Morén 2011, p. 60). Laws in the positivist sense (e.g. the fundamental laws of physics), according to critical realists, "describe how the entities comprising generative mechanisms tend to behave" (Chalmers 1988, p. 18). These laws illustrate "tendencies" which presuppose hidden generative mechanisms whose effect is to "give rise to the flux of events" (p. 18). However, such tendencies "do not in general lead to regularities at the level of events because they will typically be juxtaposed in complicated ways with other mechanisms. A falling leaf will be subject to inertial, gravitational, hydrodynamic, thermal and other mechanisms" (p. 18). According to critical realists, positivism is a form of "naïve … realism" (Gorski 2013, p. 659) in the sense that it fails to understand the real causes of phenomena, or the existence of hidden mechanisms that interact with each other. In their opinion, "the realm of the real extends beyond the realm of the senses" (Gorski 2009, p. 149). In other words, critical realists argue that positivism is mistaken because the structures of reality are not self-evident or directly observable, but only "indirectly observable" by means of their "effects" (p. 149). To observe the behavior of a falling leaf is to witness the result of a collection of complex, hidden, juxtaposed mechanisms that lie behind the world as it can be perceived via our senses.

According to critical realists, the differences between the naïve realism supported by positivism and their own view can be explained in terms of the dichotomy between open and closed systems. In their opinion, positivists (as well as fallibilists) "adhere to a closed systems ontology" (Cruickshank 2007, p. 270) in the sense that they assume that the world must be accounted for in terms of "determinism" (Lawson 2014, p. 88).[1] For critical realists, the problem with this ontology is that real world phenomena can only be

[1] Critical realists curiously consider Popper's fallibilism to be a variant of positivism. We believe that this view is questionable, but we cannot analyze in detail the reasons for our disagreement here. For an overview of the strong differences between Popper's epistemology and positivism as well as of the reasons why fallibilism is compatible with social explanations understood as explanations about open systems, see Di Iorio 2015, pp. 149–79.

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

conceived in terms of open systems, i.e. non-deterministic processes that, unlike artificial laboratory situations, are characterized by unstable and frequently changing initial conditions. Regarding an open system, the strict application of the *caeteris paribus* clause, which is required to develop detailed and deterministic predictions, is impossible (p. 88). As a consequence, if the world is composed of open systems, reality must be regarded as intrinsically indeterminist and complex. In other words, critical realists assume that science is precisely "based on gaining explanatory knowledge of causal mechanisms at work beneath the realm of observed events which are held to be open to change" (Cruickshank 2007, p. 271). According to critical realism, from the standpoint of its practitioners, science "is possible" because it "is able to explain how causal processes, which are not directly observable, operate in a stratified open system" (p. 271). A stratified open system is an open system that is characterized by various interconnected (open) subsystems. From the standpoint of critical realism, the concept of hidden causal mechanisms is thus strictly linked to the concept of stratified open systems.

Critical realists also, although for different reasons, find interpretativism unsatisfactory. They argue that it is a flawed approach because it is anti-scientific and anti-realist. They regard it as anti-scientific for three reasons. The first is that it supports a methodological dualism that denies the possibility of scientific analysis of social phenomena. Next, interpretativism is often linked to a socio-cultural relativist perspective that rejects the concept of objective truth (Gorski 2013, p. 665). And thirdly, interpretativism is a form of anti-realism or idealism insofar as it denies the objectivity of social reality and its constraints by reducing this reality to its mental interpretation. As we will make clearer in section 4, understood in these terms, interpretativism is mistaken because fundamentally, even if "social structure is not independent of human minds and activities", social reality "is independent of any particular human mind" (p. 666).[2]

According to its advocates, the originality of critical realism lies in the fact that "it shifts the emphasis from epistemology to ontology" (Cruickshank 2007, p. 264). Critical realists argue that epistemology is flawed by the "epistemic fallacy", that is "the fallacy of reducing questions about reality to questions about how we may know reality" (p. 269; see also Bhaskar (1975) [1997], p. 16). In their view, the central problem that philosophical research must solve is grasping the ontology that makes science possible. Bhaskar assumes that this ontology can be understood in terms of "transcendental deduction" from the analysis of scientific practice (see Turner 2013, p. 417). In other words, for him, the central philosophical problem is giving an answer to the following question: "what must the world be like for science to be possible?" (see Chalmers 1988, pp. 19–23). In Bhaskar's opinion, even if science is fallible and subject to revisions, it must be necessarily conceived as a knowledge about stratified open systems governed by non-deterministic and largely non-observable generative mechanisms. The ontology of critical realism is thus linked to a foundationalist perspective about the ultimate and invariable general features of reality (see Turner 2013, p. 418). This means that critical

[2] The theory of interpretation is an old and complex intellectual tradition. Different perspectives coexist within this tradition, but critical realists do not acknowledge this fact. They present the theory of interpretation in a caricaturised way. Many eminent supporters of this theory, for example, Gadamer (2002), reject both relativism and antirealism. Moreover, they develop a philosophical framework that is consistent with the fallibilist concept of unified science (see Antiseri, 2006; Di Iorio 2015, pp. 17–19; 149–79; Grondin 2002; Popper 1980).

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

realism accepts fallibilism in Popper's sense only to a limited extent because Bhaskar and his followers remove "the recognition of fallibility from the philosophical claims about ontology" (Cruickshank 2007, p. 276; see also Bhaskar 1979, p. 170; Little 2013a; 2013b; 2013c; 2013d; 2013e; 2020).[3]

3. THE NON-FOUNDATIONALISM OF ANALYTICAL SOCIOLOGY

Many of the important features which define critical realism are also inherent to analytical sociology. First, with its links to the *Verstehen* tradition and complexity studies, analytical sociology matches both critical realism's view that social systems are open and indeterminist and its assumption that what matters in science cannot be reduced to what is observable. Second, both critical realism and analytical sociology agree that social phenomena are based on generative mechanisms (see section 6). Third, in its opposition to holism understood as a theory committed to social determinism that denies the existence of universal rationality, analytical sociology rejects, like critical realism, the idea that truth is culturally relative. Fourth, as we will analyze in detail in section 5, analytical sociology is at odds with interpretativism understood as anti-realism or idealism.

The point at which these two approaches part ways is that analytical sociology, unlike critical realism, is not committed to any foundationalist claims about ontology. From the standpoint of analytical sociology, the assumption that social phenomena are indeterminist systems based on generative mechanisms is a metaphysical presupposition of research that must not be understood in terms of a strong ontological claim. As stressed by Boudon, according to the *Verstehen* tradition to which analytical sociology belongs, reality is precisely explained in terms of systems, but these systems are regarded as "forms" (in Simmel and Weber's sense), making them hypothetical and fallible mental constructs we impose on reality to give some order to it. They are not "essences", ultimate and absolute noumenal truths (Di Iorio 2015, p. 95; Di Nuoscio 2018, p. 89 ff.). Boudon (1994, p. 235 ff.) describes this fallible and non-foundationalist *a priori* perspective, which takes into account the finitude of the human mind, in terms of "revised Kantianinsm". Because of its non-foundationalist background, analytical sociology conceives the different and interconnected levels that interact within open stratified systems in terms of "levels of analysis" rather than in terms of "levels of reality" (Manzo 2010, p. 144; see also Hedström 2005, p.70). The non-foundationalist perspective of analytical sociology, which takes into account the historical variability of the metaphysical presuppositions of science, seems to us more defensible from an epistemological standpoint.[4]

---

[3] As highlighted by Little (2013c), eminent critical realists such as Elder-Vass and Hartwig interpret the ontology of critical realism in less foundationalist terms than Chalmers and Cruickshank. While Elder-Vass and Hartwig "concede that there are important passages in Bhaskar's *A Realist Theory of Science* that give an impression of aprioricity and infallibility in Bhaskar", they argue that other passages taken from Bhaskar's works show "that Bhaskar's reasoning is not aprioristic and that he regards his conclusions as being fallible and historically conditioned" (see also Archer et al. 2016). A very interesting discussion on this topic between Little, Elder-Vass and Hartwig is available on Daniel Little's blog *Understanding Society* (see Little 2013c). Like Chalmers and Cruickshank, Little considers Bhaskar to be a foundationalist (see Little 2013a, 2013b, 2013d) and defends a more "fallibilist" and "anti-foundationalist" perspective in ontology and epistemology (Little 2020, p. 1).

[4] For further criticisms of the ontological assumptions of critical realism, see Mäki (2011) and Mäki and Oinas (2004).

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

Chalmers (1988, p. 19) explains the shortcomings of critical realists' foundationalism in the following terms:

> Suppose a medieval scientist, working within some version of Aristotelian theory, had asked Bhaskar's question "what must the world be like for science to be possible?" and had attempted to answer it in a Bhaskarian way. I suggest a most plausible part of his answer would have been "the world must be a finite, harmonious whole with a centre". The whole of Aristotelian science, with its distinctions between forced and natural motions, terrestrial and celestial regions and so on, presupposed the absolute space provided by such a cosmos. This is the conclusion that a medieval version of Roy Bhaskar would have arrived at. We now know, of course, that that conclusion was fallacious. In retrospect we can see that Aristotelian physics did presuppose a cosmos. We can also understand why that approach would have limited success in spite of the fact that it involved false assertions and has been replaced by a more successful theory and practice. It is not difficult to imagine our current science suffering a similar fate. The practices that Bhaskar correctly identifies as important components of contemporary science may be transcended by more successful practices. In that case a future historian might acknowledge that Bhaskar had correctly identified a world view presupposed by the science of his day but one that proved to be of limited applicability and inadequately geared to coming to grips with the world as it really is.

4. FOUR CRITICISMS OF METHODOLOGICAL INDIVIDUALISM

A major flaw of analytical sociology, at least from the perspective of critical realists, is that this approach is committed to methodological individualism (Gorsky 2009). In their opinion, methodological individualism is mistaken for four reasons. The first is that it is a form of "social atomism" (Bhaskar 1979, p. 29; see also Archer 1995, p. 35). As such, methodological individualism is reductionist in the sense that it assumes "that facts about societies, and social phenomena generally, are to be explained solely in terms of facts about individuals" (Bhaskar 1979, p. 29). Critical realists draw this reductionist interpretation of MI, which is rooted in Marx's and Durkheim's works, from analytic philosophers such as, for example, Lukes (1968), who reframed the old view that MI equates with social atomism in terms of language analysis (see Archer 1995, p. 31 ff.). On this reading, MI assumes that, to develop satisfactory accounts of social phenomena, all statements about social wholes (e.g. states, societies, armies, economies, religious sects and so on) must be reduced to statements about the particular individuals comprising them (see Bhaskar 1979, p. 29); or, if one prefers, that all social predicates must be semantically reduced to individual predicates. Like many analytical philosophers, critical realists argue that MI must be rejected because there are social properties that cannot be semantically reduced to strictly individual proprieties: the whole is greater than the sum of its parts. As a consequence, facts about societies cannot be explained solely in terms of facts about individuals. If, for example, we compare the economic systems of two countries and argue that "country X is richer than nation Y", we use a sentence that cannot be semantically reduced to a set of predicates concerning properties of individuals because this sentence takes into account statistical aggregates that refer to global properties and thus does not entail that any member of country X is richer than any member of country Y. Understood as a reductionist approach, MI is regarded by critical realists as incompatible with concepts such as "emergence", "structure" and "open systems" that, in their view, are "quintessential" to any correct scientific analysis of the social world (Archer 1998, p. 190).

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

Critical realists' second motivation for rejecting MI is that they consider it to be a form of "voluntarism" (Bhaskar 1979, p. 34), by which they mean an approach that regards society as created by "meaningful" human intentions that are not subject to social conditioning. In their opinion, assuming the existence of society as depending on the "unconditioned" (p. 34) will of agents is another aspect of the individualist commitment to atomism. Critical realists do not criticize MI's voluntarism in the name of a defense of holism in Durkheim's sense. According to them, the latter is equally mistaken insofar as it hypostatizes society and argues that enduring human relationships are created by a substantial entity that exists independently of human agency (see Archer 1998, p. 192; Archer, 1995, p. 31 ff.). In critical realists' view, the connection between society and the conscious activity of people can be explained neither in terms of MI's voluntarist upward causation, nor in terms of holism's determinist top down causation, but depends on a circular causality (feedback loop) between agency and preexisting social forms:

> people do not create society. For it always pre-exists them and is a necessary condition for their activity. Rather, society must be regarded as an ensemble of structures, practices and conventions which individuals reproduce or transform, but which would not exist unless they did so. Society does not exist independently of human activity (the error of reification). But it is not the product of it (the error of voluntarism). (Bhaskar 1979, p. 39)

Social dynamics are based on a structure of "persistent relations" (p. 31). Moreover, these dynamics are also influenced by other aspects of social conditioning such as socialization. As a consequence, social dynamics stem from a dualism between intentionality and its "antecedent conditions" (p. 39). Actions, including those that aim at transforming society, are neither mechanically determined by holistic factors, nor completely free. Bhaskar (p. 39) clarifies how individual intentionality is socially influenced, taking into account the relationship between language and intentionality:

> we can allow that speech is governed by the rules of grammar without supposing either that these rules exist independently of usage (reification) or that they determine what we say. The rules of grammar, like natural structures, impose *limits* on the speech acts we can perform, but they do not *determine* our performances. This conception thus preserves the status of human agency, while doing away with the myth of creation (logical or historical), which depends upon the possibility of an individualist reduction.

The "pre-existence" of the social forms, which are "a necessary condition for any intentional act", establishes "their autonomy as possible objects of scientific investigation" (Bhaskar 1979, p. 25). According to Bhaskar, due to its atomistic and voluntaristic assumptions, MI cannot take into account this autonomy. Rejecting MI, critical realism assumes that "the social context has autonomy and independence from people, pre-exists them, and is causally influential of them" (Archer 1995, p. 46). As opposed to the creationist approach of MI, critical realists' explanatory model is called by Bhaskar (1979, p. 39) a "transformational model of the society/person connection". This model is partly inspired by Marx's view that historical phenomena must be studied in terms of pre-existing dichotomic relations structuring human agency such as, for instance, the capitalists-wage-workers relation. However, critical realists reject the more holistic aspects of Marx's approach, namely his economic determinism: "material production is a necessary condition for social life", but "it cannot be proved that it is the ultimately determining one" (p. 45).

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

For Bhaskar (p. 38), the transformational model of social explanation that is supported by critical realism must be conceived as linked to the theory of the unintended consequences of human action:

> people, in their conscious activity, for the most part unconsciously reproduce (and occasionally transform) the structures governing their substantive activities of production. Thus people do not marry to reproduce the nuclear family or work to sustain the capitalist economy. Yet it is nevertheless the unintended consequence (and inexorable result) of, as it is also a necessary condition for, their activity. Moreover, when social forms change, the explanation will not normally lie in the desires of agents to change them that way, though as a very important theoretical and political limit it *may* do so.

The third reason why critical realists reject MI is its connection with the *Verstehen* tradition. They accuse MI of being a form of interpretativism, an anti-realist or idealist approach that denies the objectivity of social reality by reducing this reality and its constraints simply to mental constructs (see Di Iorio and Chen 2019; King 2004, pp. 68–72; Turner 2013, p. 418). According to critical realists, MI cannot account for the fact that the social world and its bounds exist independently of the individual interpretation of this world, that is, independently of the individual's opinion about what they are free or not free to do (see King 2004; Di Iorio 2016, p. 7). In their opinion, for example, it is precisely because legal restrictions are not subjective, but objective, that ignorance of the law does not make a person free to transgress them without consequences. As stressed by Gorski (2013, p. 666), the structure of social reality and its constraints "is independent of any particular human mind" (see also Lawson 1997, p. 137; Bhaskar 1979, p. 28 ff.). Since MI does not define social forms "in objective terms … but only … in terms of human beliefs" (Lawson 1997, pp. 138–9), it lacks an "adequate … ontology" (p. 148) of the social world. In other words, since MI assumes that society consists "ultimately, in the opinions, beliefs and attitudes of individual agents" (p. 139), it is a psychological reductionism that mistakes the social with the mental. To use Archer's words, from the standpoint of MI, "the relevant facts are … mentalistic" (Archer 1995 p. 36). Critical realists argue that the "structures" that govern and influence human interactions are "irreducible to individual conceptions" (Lawson 1997, p. 139) even if they "are dependent upon … the concepts and actions of human agents" (p. 147). There exists an objective reality that is binding on the agent that must be explained in systemic terms and consists of "social practices", "social rules" and "social positions" (Lawson 2003, p. 39).

The fourth and last reason why critical realists reject MI is because the latter considers rationality to be the key to explaining social phenomena. By rationality critical realists mean "a … theory of efficient action" that is strictly instrumental and utilitarian in the sense of the rational choice model (RCM) (Bhaskar 1979, p. 32; see also Archer 1995, pp. 210–11). According to critical realists, understood in these terms, rationality is "an *a priori* presupposition of investigation, devoid of explanatory content and almost certainly false" (Bhaskar 1979, p. 32). This is so for two reasons. The first is that it is problematic to assume that the human mind works in a strictly instrumental and utilitarian way (Archer 1995, pp. 210–11; Little 2018). The second is that arguing that agency is based on rationality means neglecting that agency does not depend only on intentionality, but also on the structure of the social relations which shape it (Archer 1995, pp. 210–11). Because of this, even when there is room to describe social phenomena in utilitarian and instrumental terms, a relevant point is missed: "to say that people are rational does not explain what they do, but only at best … how they do it" (Bhaskar, 1979, p. 32).

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

5. IN DEFENSE OF NON-ATOMISTIC METHODOLOGICAL INDIVIDUALISM

Critical realists' objections to MI fall down in their reliance on a reductionist account of the latter which incorrectly commits it to social atomism. In fact, social atomism is only the most simplistic variant of MI. The criticisms leveled at MI by critical realists are based on a caricaturised interpretation of this approach. Analytical sociology rejects the atomistic variant of MI used in mainstream economics in preference for a non-atomistic variant that can be called "complex individualism" (Manzo 2007, p. 4) or "structural individualism" (Bearman and Hedström 2011, p. 4). This non-atomistic variant is rooted in the tradition of interpretative sociology, as well as in the works of Popper and the Austrian school of economics (see Demeulenaere 2011, p. 12; Di Iorio and Chen 2019).

Reply to criticism 1. Both analytical sociologists and their precursors agree on the groundlessness of the reductionist and non-systemic view that all statements about social properties must be replaced by statements about individuals (see, for example, Popper 1957, p. 82). Given the above, as argued by Manzo (2014, p. 21), analytical sociology's non-atomistic approach "can handle: (1) macro- and meso-level entities and properties; (2) downward causation; (3) macro-level emergence; and (4) macro-to-macro and meso-to-meso correlations". The confusion between MI and semantic reductionism promoted by analytical philosophy—and, in its wake, by critical realism—is linked to a misunderstanding of the individualist claim that social explanations must be "in terms of individuals" (Kincaid 2017, p. 87; see also Bhaskar (1975) [1997], p. 29). By explanations of this kind, MI, or at least its non-atomistic variant, does not mean non-systemic explanations in the sense of reductionism as understood, but non-holistic explanations that reject socio-cultural determinism because they attach central relevance to human action (for more details on this point see Di Iorio 2015, pp. 75–120). Moreover, the commitment of MI to the study of the unintended consequences of human action, in other words unforeseen systemic effects, makes clear that, according to this approach, the vocabulary of social properties is not semantically reducible to the vocabulary of individual mental and behavioral properties (see Di Iorio and Chen 2019, p. 361; see also Bouvier 2011; Bulle 2019; Linares 2018).

Reply to criticism 2. Non-atomistic MI presupposes that, since human interactions are influenced by preexisting social factors and "structures of … relationships" (Hayek (1952) [1948], p. 39), society is not explainable in terms of "voluntarism" (Bhaskar 1979, p. 34; see also Hayek (1952) [1948], pp. 1–32; Boudon 1971, 2013). According to non-atomistic MI, "individual action always occurs within the framework of a system of constraints which are more or less clearly defined, more or less transparent to the subject, and more or less rigorous" (Boudon and Bourricaud 1990, p. 13). Moreover, like critical realism, non-atomistic MI assumes that social order and social conditioning largely result from unintended consequences that are related to shared and historically conditioned normative beliefs. Making jokes about a person dressed in a manner that we consider ridiculous ensures, for example, social control, even if ensuring this control is not our goal (see Di Iorio 2015, p. 99). The difference between non-atomistic MI and holism is that, even if the former conceives social phenomena in terms of "interdependent elements" (Boudon and Bourricaud 1990, p. 395) and structural constraints and social positions, it does not locate the causes of action in these social factors (Dawe 1970, pp. 209–11). This is because non-atomistic MI, which acknowledges human autonomy, regards these factors as the "parameters" of action, not as the "causes" of action (Boudon 2013, p. 30).

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

In addition, this approach rejects the idea that social constraints are mechanical and affect the individual independently of their lived experience (see Crozier and Friedberg 1980, p. 17 ff.). It seems clear to us that analytical sociology and critical realism agree not only on the fallacy of holism, but also on the fallacy of voluntarism and social atomism (see also Bulle and Phan 2017; Chen 2020).

Reply to criticism 3. The assumption that MI is mistaken because social constraints are objective in the sense that they exist independently of the agent's opinion about what they are free or not free to do is based on a misunderstanding of the *Verstehen* approach. This is because, contrary to what critical realists argue, according to MI and especially its sociological and non-atomistic variant, strictly subjective opinions "are not the basis of social life" (King 2004, p. 190). In explaining the social world and its constraints, MI does not apply the *Verstehen* approach to personal opinions, but rather to "collective beliefs" (Boudon 2001), to typical ways of thinking and feeling (Weber 1978). MI describes social constraints and social sanctions as objective and real consequences of these shared beliefs—consequences which are often unintentional. As highlighted by Hayek ((1952) [1948], p. 34), social structures must be regarded as "the implications of many people holding certain views", i.e. as the product of "a mental structure common to all of them". The simple fact that, in explaining social phenomena and structural constraints, MI conceives the consequences of shared beliefs as often unintentional shows that this approach is at odds with psychological reductionism and anti-realism. Given the commitment of MI to the study of unintended consequences, the orientation of this approach must not be confused with the reduction of the social to the mental (see Di Iorio 2016). As pointed out by Popper (1963. pp. 124–5), according to MI, "the real task of the social sciences is to explain those things which nobody wants".

Reply to criticism 4. The association between the non-atomistic variant of MI and the RCM "is unfounded" because the former "is neutral with respect to the type of theory of action chosen to portray actors' micro-level activities" (Manzo 2014, p. 21). Analytical sociology does not assume that rationality is always instrumental and necessarily based on the attempt to maximize subjective utility. Within non-atomistic MI, different theories of rationality coexist (Boudon 1995, p. 251). As argued by Boudon (2007, 2009, 2013), this variant of MI is compatible with a concept of rationality that is broader than the one used by the RCM (see also Demeulenaere 2011 pp. 26–7; Manzo 2013a, pp. 373–4). This broader concept of rationality, which is called *ordinary* or *cognitive* rationality (Demeulenaere 2011 pp. 26–7), and is rooted in the theory of argumentation, allows us to understand both "positive beliefs (i.e. descriptions of the world and its mechanisms) and normative beliefs (i.e. ethical beliefs) that are the outcomes neither of a socio-cultural determinism nor of a utilitarian motivation" (Di Iorio 2015, p. 122). For more details on this point see Boudon (1994, 2001) and Bronner and Di Iorio (2018). It must be added that, according to a number of analytical sociologists, non-atomistic MI is compatible not only with different theories of rationality, but also with different theories of action (Manzo 2014, p. 22). This is because these scholars assume that MI is not necessarily connected to the rationality principle and highlight the need to explain action by taking into account, for example, insights from biology and evolutionary psychology (León-Medina, 2014, 2017; Lizón 2016).

According to non-atomistic MI, the existence of individual rationality is compatible with the existence of social conditioning. We have already pointed out that, according to this approach, actions are not "'the product of a 'free agency' or of an individual liberty which is conceived

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

as absolute" (Boudon and Bourricaud 1990, p. 13). From the standpoint of non-atomistic MI, the agent "operates in a framework of constraints determined both by socialization and by the structure of the situation" (p. 14). As a consequence, understanding their rationality means taking into account this framework and the way it affected their action (pp. 287–9). According to non-atomistic MI, "the idea of rationality must thus be seen as relative, that is to say as dependent upon the structure of situations. To be sure, it must also be seen as dependent upon the position and generally the characteristics of the actors" (p. 290).

6. SOCIAL MECHANISMS AND CAUSALITY

The origins of the critical realist conception of mechanisms can be found in Rom Harré's *The Philosophies of Science* (1972). According to Harré (1972, p. 170), explaining means identifying the underlying generative mechanisms of individual happenings or sequences of events. For him, a generative mechanism is the way a thing acts in response to a certain stimulus *by virtue of its nature*. This conception of mechanisms is thus necessarily linked to a reflection on the nature and the constitution of the thing that causes the effect (1972, p. 124). Bhaskar, who wrote his thesis under the supervision and influence of Harré, also conceives of mechanisms in a way that is marked by the idea of identifying the potentialities that lie in the nature and inner constitution of those entities that we consider as causes. In consequence of Bhaskar's own influence, critical realists conceive of mechanisms as *causal power*s or *liabilities* of objects or relations (Sayer 1984 [1992], pp. 104–5). A mechanism is what an object can or tends to do or, in other words, "the ways of acting of things" (Bhaskar (1975) [1997], p. 3) or how objects "work" (Danermark et al. 1997 [2005], p. 54). Those causal powers exist by virtue of the object's nature (Sayer 1984 [1992], p. 105), and this nature is said to consist basically of its inner structure (Collier 1994, p. 43). In this section, we discuss why we think this approach is flawed, especially in its application to the social sciences.

6.1 Emergence

The critical realist conception of mechanisms and causation can be usefully illustrated by placing alongside each other the three claims that it involves: *Claim 1*: when a set of entities is complexly structured a new ontological higher-level entity is formed; *Claim 2*: every higher-level entity has its own properties that are different from those of its components, so those properties cannot be explained reductively; *Claim 3*: the inner structure of an entity is causally efficacious, and that includes downward causation: original, lower-level entities acquire new causal powers and properties that are caused by the inner structure of the new higher-level entity.

Comments on claim 1. The first claim (macro-level entities have ontological status) is far from being obvious, even in the fields of physical and natural sciences, but the observable nature of entities such as atoms, molecules or individuals means that, for the purposes of scientific inquiry, there is not much difference between stating their ontological existence and just considering them as if they were real. However, while atoms or individuals are directly observable, social institutions are not, so the unobservable nature of social entities adds a new difficulty. Critical realists should be able to offer very powerful reasons for us to believe that

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

unobservable entities really exist, but the truth is that they fail to do so. The argument that they exist because they are causally efficacious (Bhaskar 1998 [1979], p. 27), known as the *causal criterion of reality*, has already been convincingly refuted by Hedström (2005, p. 72): to believe in the existence of causally efficacious unobservable entities we need something more than a reference to an alleged existence of causal powers that could remain always inactive (and therefore unobserved), or whose power could be counteracted by other causal powers (with no methodology to isolate them and therefore determine their existence).

Comments on claim 2. Critical realists also state that complexly structured entities have their own properties that cannot be reductively explained. The main argument here is that collective properties cannot be found in the constituents' set of properties (Bhaskar 1998 [1979], p. 38; Elder-Vass 2005; Gorski 2009, p. 157), so they can be said to be "emergent". From our point of view, to state that social collectives (whether we consider them really existing entities or not) have properties that are not present at the individual level is a truism, so analytical sociology accepts, as we have already said in section 5, that collective properties cannot be replaced by statements about individuals. But this in no way means that collective properties cannot be explained, in a bottom-up fashion, as the result of complex dynamics of interaction between micro-level entities. The problem with Critical Realism here is that it adheres to an interpretation of emergence that unavoidably leads to stating the existence of a "mystery gap" between micro and macro levels, which is not a very "realistic" approach. As we understand it, the long tradition of the microfoundation strategy, to which analytical sociology belongs, aims to show empirically, case by case, that more or less complex processes of aggregation are always behind the emergence of any systemic property.

Comments on claim 3. At least in the social sciences, it is quite different to state that systems have their own properties and to state that they have autonomous powers, as critical realists do in their third claim. Needless to say, critical realists are not alone in this. The idea that only social facts can explain social facts is a foundational myth of sociology, and the idea of social forces shaping human behavior is also mainstream in our field. Among critical realists, this is settled on the claim that the inner structure of a social object is causally efficient, even in a downward fashion. Analytical sociology, however, has always stated that considering social structures as entities with their own autonomous causal powers introduces an element of unintelligibility to sociological explanations. Intelligibility requires a causal agent who can bring about change. Social conditions, systems' properties, structural constraints etc. cannot bring about change, so they cannot be considered as causal agents. Accepting an autonomous social force as a proper *explanans* precludes the construction of final black-box-free explanations.

As we stated in section 5, denying that social structures have autonomous causal powers does not imply that bottom-up explanatory strategies should dismiss the role of social structures or systems' properties. Quite the contrary, structural constraints should be taken into consideration, either to understand agents' behavior and interactions, or to explain why behavior and interactions produce certain outputs. However, from an analytical point of view: (1) they should always be thought of as effects of more or less complex processes of action and interaction. An example will clarify what we mean: the set of links within a social network, and all its properties, can be key elements in understanding individual decisions and how they complexly aggregate to form a social phenomenon, but networks are nothing but formal representations of human activities (A knows B, A hates B, A is loved by B …).

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

The fact that we use them as elements of an explanation does not mean that we do not recognize that they can be shown to result from actions and interactions. Furthermore (2), they should enter our explanation as parameters and not as causes of individual behavior, and/or as necessary conditions for a certain aggregation process, not as efficient causes of that process.

In considering downward causation, critical realists insist that some types of relation cause new causal powers in the individuals involved. For example, in the landlord-tenant relation, the landlord's powers cannot be explained as the result of his individual properties (Danermark et al. 1997 [2005], p. 54), but instead as an effect of the relation itself. The same would remain true of other relations, such as those between capitalist and worker, husband and wife, or any other *substantial*, or *internal and necessary* relation (Sayer 1984 [1992], p. 88–9). Critical realists are therefore pretending that the existence of relational properties (properties that I only have as long as I am in a certain type of relation) posit a challenge for bottom-up explanatory strategies and that they are proof of the causal efficacy of social structures. This is repeatedly stated and never justified. The truth is that the non-atomistic MI to which analytical sociology adheres (see section 5) does not rule out relational properties, as different authors have already stated (see, for example, Boudon 1986, p. 55; Noguera 2003, p. 107).

In sum, the critical realist statement according to which collective properties have "causal powers" is only acceptable for analytical sociologists if it means that collective properties can be included in a sociological explanation as necessary conditions for an effect to be produced, but not if it means that they can be considered as efficient causes. Since the only efficient cause of social phenomena is human action, assuming collective properties to be efficient causes would mean introducing an unintelligible element into sociological explanations and, therefore, an undesirable black box. An example chosen by Sayer will clarify the distinction between the critical realist and the analytical explanatory strategy. According to him, "… the (approximately) regular flux of traffic in and out of the city is an effect of the rules governing the working day and the separation of home and workplace" (1984 [1992], p. 124). A bottom-up strategy for explaining the regular flux of traffic in and out of the city would state that nothing but the decision and the actual behavior of driving at certain moments of the day can be regarded as efficient causes of the *explanandum*. Only individual actions change the scene. The separation of home and workplace does not cause anything, even when we agree that without that separation we would observe a different pattern in the flux of traffic. But necessary conditions cannot be regarded as causes. Moreover, the rules governing the working day, though indispensable to explaining the regular flux of traffic, are not existing entities that could efficiently cause it. Social rules are nothing but shared agreements that can be easily shown as a form of *collective intentionality* (mental states that I have as long as others have them too). They have no existence outside humans' brains (even if they are objective and real in the sense clarified in section 5). Those rules are parameters of the individual decision, but not causes, as we stated in section 5 (see also Boudon 2013, p. 30). To sum up, when Sayer presents the separation of home and work and the rules governing the working day as necessary and sufficient explanatory elements of the regular flux in and out of the city, he is accepting a superficial black-box explanation of a phenomenon that is easy to microfound. Analytical sociology operates on the principle that any explanation of social facts or individual attributes as effects of other social facts can be broken down to a microfounded black box–free explanation.

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

6.2 Mechanisms, Regularities and Generalizations

Critical realism describes a reality composed of three different domains: the empirical, the actual and the real (Bhaskar (1975) [1997]). The empirical is the part of the world that we can have access to and what we can experience (our data, the fact that we describe). The empirical is just a subset of the actual domain, where events take place whether we observe them and know them or not. These events are produced in a deeper and barely accessible domain where mechanisms are located: the real domain. Thus, mechanisms are conceived as causal powers that crouch behind the empirical, shaping them from a hardly observable position. They can exist but remain untriggered (and therefore unobserved), and they can be triggered but still remain unobservable. In the latter case, they can fail to produce their effects because other mechanisms can counteract them. Mechanisms are causal powers that generate events, but they are not particular sequences of events, they do not trigger any particular sequence of events and their triggering conditions are usually only identifiable as such *ex post* (Danermark et al. 1997 [2005], pp. 55–6; Sayer 1984 [1992], pp. 104–5).

We would like to draw attention to two important consequences of this conception of mechanisms. The first is that mechanisms research becomes nothing more than a speculation about invisible structuring forces of reality. Even if we pompously call it *transcendental argumentation*, the truth is that these philosophical operations are presented as free of any obligation of being subject to empirical verification. Empirical data has nothing to say about mechanisms thus conceived. We shall develop this idea in the seventh section of this chapter.

The second consequence is that sociological research would be doomed to producing only idiographic knowledge. Each phenomenon would be nothing but a unique combination of specific activators triggering previously dormant mechanisms that combine in unpredictable ways, producing unpredictable consequences. The possibility of finding regularities in the observation of different cases of *x* that could lead us to legitimate generalizations is ruled out, at least in the traditional sense of generalization, which is disregarded as "empiricist". Danermark et al. (1997 [2005], p. 76), for example, oppose this "empiricist" concept of generalization (understood as generalization to a greater population or set of phenomena) and the "critical realist" concept (understood as a thought operation to identify *transfactual conditions*, that is, more or less general conditions for an object to exist). The former would only capture the superficial empirical domain, while the latter is supposed to have access to the deep real domain.

Like critical realists, analytical sociologists propose a thorough rethinking of the role of statistical regularities and generalizations in the construction of sociological explanations. However, while analytical sociologists state that statistical regularities *alone* cannot say anything about causality, critical realists state that statistical regularities cannot say anything *at all* about causality. The analytical conception of mechanisms is compatible with a recognition of the valuable role of quantitative research precisely for its capacity for grounding the construction of legitimate generalizations in the observation of statistical regularities (Hedström 2005, p. 101). Even more importantly, it does not rule out the possibility of identifying regularities in the activation conditions of mechanisms and in the consequences of those activations. In fact, in the analytical tradition, social mechanisms are always conceived as *regular* ways in which *typical* sets of linked individuals produce different *types* of social *regularities* (see, for example, the definitions of Hedström 2005, p. 33 or Manzo 2014, p. 7). In this way, mechanisms research does not give up generalizability as an essential scientific task. It simply reformulates its role in the construction of satisfactory explanations.

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

7. EXPERIMENTS, SIMULATION AND THE SOCIAL SCIENCES

One of the most important consequences of the tripartite "ontological map" that divides reality into empirical, actual and real domains is the statement that science should not be limited to a passive observation of events and the identification of statistical patterns and regularities. For critical realists, science is not about describing what we observe in the empirical domain, but about having access to the real domain. However, having access to this deep and not immediately observable domain requires an active and systematic intervention. From this point of view, critical realists state that experiments are the only way of isolating a mechanism to verify whether it exists and to determine its effects (Bhaskar 2009 [1986], p. 70; Collier 1994; Manicas 2006).

While critical realists see experiments as a privileged tool for having access to the real domain, they also claim that experiments cannot be conducted in the social sciences. An alleged impossibility of achieving the necessary closure for isolating social mechanisms is presented as a definitive argument for dismissing the method (see, for example, Collier 1994; Danermark et al. 1997 [2005]; Gorski 2013; Sayer 1984 [1992]). In contrast, analytical sociologists tend to consider that experimentalists have managed the difficulty of achieving the necessary closure reasonably well. Analytical sociology's objection to the experimental method does not focus on whether it is possible, but on its limits and, more specifically, its theoretical relevance for an action-based approach (see, for example, Goldthorpe 2001 or Hedström 2005, pp. 103–4).

What primarily separates analytical sociology and critical realism, though, is not their respective motivations for skepticism about the applicability of experiments in the social sciences, but the ways they attempt to circumvent the perceived limitations. What is the alternative to isolating mechanisms in the social sciences? Is there a way to study how different social mechanisms combine to produce what we observe?

The unsolvable distance between analytical and critical realist methodologies is better understood if we take into consideration the distinction between *retroduction* and *retrodiction*. The former refers to the process identifying a mechanism, while the latter refers to the concrete research that examines how mechanisms combine to cause particular events (Elder-Vass 2007, p. 465; Sayer 1984 [1992], p. 116). As we will see, while it is true that analytical sociologists and critical realists share a common emphasis on the central role of these modes of inference, they do not share much more than this.

7.1 Retroduction

Social phenomena always occur in open systems where different mechanisms, when activated, can cooperate, reinforce, complement, counter or even cancel out each other, with no possibility for scientists to use empirical data to isolate them and determine their existence. Critical realists try to compensate for the alleged impossibility of experiments by turning their attention to conceptual abstraction, which is regarded by some authors as "the most central social scientific activity" (Danermark et al. 1997 [2005], p. 41). Abstract concepts are the result of a thought process where we identify the indispensable and constitutive powers and properties of something and remove the less important and contingent parts. In fact, for critical realists, causal processes are something that we can only grasp by reflecting on the inner nature of the

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

object or, in other words, constructing *transcendental arguments* (arguments in which we ask what must be true in order for x to exist?).

In the words of Danermark et al.:

> … conceptual abstraction is used as a kind of social science equivalent to the natural science experiment. […] when we wish to gain knowledge about generative powers and mechanisms in social worlds, one of the most splendid tools at our disposal is the isolation of certain aspects in thought—abstraction—rather than isolating them by manipulating events. (1997 [2005], p. 43).

Knowing the causal powers of an object requires, according to Bhaskar, "retroductive arguments" (arguments that move "from a description of some phenomenon to a description of something which produces it or is a condition for it" (2009 [1986], p. 7). A critical realist is then someone "who claims that […] mechanisms can be shown to be real by means of transcendental arguments" (Collier 1994, p. 26).

Empirical research offers nothing to the knowledge of social mechanisms thus conceived. For critical realists, what we have in the empirical domain is a mess of data irrelevant to identifying causal processes. The causal powers of an object result from its inner structure (Collier 1994, pp. 7–8), a structure that has nothing to do with the appearances that we observe in the empirical domain. This lack of concern for empirically testing theories is exemplified in Sayer's reference to Marx's mechanism for the "tendency of the rate of profit to fall". Sayer states that this mechanism cannot be considered false just because one always fails to observe it. In the end, other counter mechanisms could be responsible for persistently cancelling this tendency. If we consider it false, he argues, it is not because it lacks empirical support, but because Marx's reasoning on the nature of capital is flawed (1984 [1992], p. 106). In this way, considering the identification of mechanisms as a mere conceptual task, critical realists dismiss the role of empirical tests in the construction of scientific knowledge (see, for example, Danermark et al. 1997 [2005], pp. 135–6).

Two features of the analytical approach lead to a very different conception of retroduction. First, mechanisms are conceived by analytical sociologists as micro-level causal processes that bring about macro properties or phenomena. And second, analytical sociologists accept Merton's notion of middle-range theories (Hedström and Udehn 2009), especially regarding how to determine a proper *explanandum*. Taken together, these two features have led analytical sociologists to conceive retroduction as a process of determining whether a certain micro-level mechanism is sufficient to generate a particular social phenomenon. Computer simulations, more specifically agent-based models, have been used in the analytical community as a way to achieve the difficult isolation of social mechanisms. In Hedström's words, "Analyses using [agent-based] models should be viewed as virtual laboratory experiments; they assume a closed social system in order to establish the causal tendencies of different social mechanisms." (2005, p. 99). A practical example can be found in Manzo's simulation on the reproduction of educational inequalities (2013b), a simulation that was built to isolate and assess the relative explanatory contribution of two different mechanisms that could potentially explain educational choices: ability/benefit evaluations and peer-group pressures.

This process of determining the generative sufficiency of a mechanism is neither entirely conceptual nor entirely empirical. On the one hand, conceptual abstractions have a role, but they are seen as empirically founded constructions that help us build formal models (to be subsequently "translated" to computer models) that have no other ultimate purpose than

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

being empirically tested. Concepts are grounded in empirical data and are intended to form hypotheses to be tested with more empirical data. Moreover, they refer to actual but stylized traits of objects and relations, not to their inner and unobservable potentialities or "essential" properties. On the other hand, empirical work is also necessary to calibrate the model settings and parameters so we can claim to be generating a replica of the natural process in silico. When the model is empirically calibrated, as in Manzo's previously cited model (2013b), what we run in our computers is not just a thought experiment extended with the aid of technology, but an empirical test in itself.

7.2 Retrodiction

For critical realists, conceptualization lays the foundation for empirical research, where scientists analyze how different mechanisms interact in a particular situation. If retroduction is principally a thought exercise, retrodiction is empirical by nature. However, critical realism can only be vague on this point. Recommendations of social research techniques for critical realists are limited to the easy appeal of combining qualitative and quantitative techniques, or combining intensive and extensive research (Danermark et al. 1997 [2005]). In the end, qualitative/intensive research designs are seen as the only way of producing causal explanations and quantitative/extensive designs are seen as just an aid in the description and a source of suggestions for possible mechanisms at work.

Against this somewhat slipshod application of retrodiction a series of objections can be elicited from an analytical point of view. The first is that, no matter how hard qualitativists have tried, qualitative research has not yet resolved the problem of inference (Goldthorpe 2000), and we do not want to diminish the social sciences to the task of building sophisticated descriptions of very particular events. Instead, analytical sociology turns its attention to middle-range theories that aim to create empirically validated statements with a certain degree of generalization. Next, when retroduction has been reduced to a thought operation, proposed mechanisms with no empirical evidence whatsoever might be left in the position of structuring and conditioning research. The possible consequence of research being misleading and/or ideologically biased is quite obvious. Third, happily, the social sciences in the twenty-first century are no longer limited to qualitative/intensive research when they aim at describing processes of overdetermination. As analytical sociologists already know, there is a powerful tool for analyzing how different mechanisms interact, which can produce statements of their combined effect whose generalizability may be empirically tested later on. One would expect that a philosophical trend that emphasizes the relevance of retrodiction in the construction of causal explanations would have developed an interest in agent-based models. However, only a few attempts to take them into consideration have been made by critical realists (see, for example, Miller 2015; Törnberg 2019), and these attempts, especially that made by Törnberg, proposes the use of simulations only for retroduction (as a tool for conceptual abstraction) and not for retrodiction (therefore stating that models do not need to be empirically calibrated). Critical realists' general oversight of computer simulations is not just a negligible oversight, however, but the result of an epistemological incompatibility. Agent-based models are not just tools for conducting research. Their use also invites us to make important reflections on the epistemology and ontology of social facts (see, for example, Humphreys 2009 and Manzo 2020). Among them, we would like to highlight two with radical consequences for the critical realist approach. First, nothing happens in an agent-based model if you do not press the "go" button, as nothing would happen

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

in society if we could press "pause" and freeze all individuals (Hedström 2005, p. 29); that is, all social dynamics and processes are the product of agents' actions and interactions. Structural constraints can be there conditioning how agents behave, or even conditioning the result of agents' behavior, but structures do not cause events and therefore cannot explain social change. And second, there is no social phenomenon that could not be reproduced as the consequence of agents' actions and interactions in an agent-based model. The possibility of "translating" a concept that aspires to capturing some relevant social trait to the language of an agent-based model can even be considered as a yardstick of its clarity and precision. If, for example, we cannot build a formal model that includes the phenomenon of the *habitus*, it would not be due to any essential impossibility of the phenomenon to be formalized, but as a consequence of how blurry the concept is. Give a modeler a clear and precise sociological concept and they will reproduce it in an agent-based model, either as part of the initial setting or as one of its outputs. The epistemological consequences of these models are clear: there is no property of social systems that cannot be shown to be a result of structurally conditioned actions and interactions of individual agents. There is no mystery gap between the micro and the macro levels.

8. CONCLUSION

Our purpose has been to compare the perspective of analytical sociology to that of critical realism. In addition, we discussed some shortcomings of the way the latter approach conceives the nature and presuppositions of science. Critical realists have levelled various criticisms against analytical sociology. We have analyzed these criticisms in detail and attempted to dismiss them with the aim of defending the research program of analytical sociology. Our line of reasoning can be summarized as follows:

(i) Critical realism and analytical sociology agree on some common points, namely the idea that social systems are open and indeterminist, the inadequacy of both positivism and relativism, the fact that social phenomena are based on generative mechanisms and the limits of both holism and atomism.
(ii) A crucial difference between critical realism and analytical sociology is that the former is based on foundationalist ontological assumptions about the deep structures of reality, while the latter is not.
(iii) The non-foundationalist perspective of analytical sociology, which takes into account the historical variability of the metaphysical presuppositions of science, is more defensible from an epistemological standpoint than Bhaskar's foundationalism;
(iv) Critical realists criticize analytical sociology because of its commitment to methodological individualism (MI). This is because they assume that MI is tantamount to social atomism.
(v) Analytical sociology's commitment to MI does not entail that the former is tantamount to social atomism because social atomism is only the most simplistic variant of MI. Analytical sociology is based on a non-atomistic variant of MI that can be called "complex individualism" (Manzo 2007, p. 4) or "structural individualism" (Bearman and Hedström 2011, p. 4).
(vi) Critical realists rightly state that collective properties cannot be found in the constituents' set of properties, but they wrongly conclude that collective properties cannot be explained in a bottom-up fashion.

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

(vii) From an analytical point of view, considering social structures as entities with their own autonomous causal powers, as critical realists do, introduces an element of unintelligibility in sociological explanations.
(viii) Critical realists conceive mechanisms as causal powers that crouch behind the empirical domain. Empirical data has nothing to say about mechanisms thus conceived and mechanisms research becomes nothing but a speculation about the inner nature and constitution of invisible structuring forces of reality.
(ix) Critical realists state that experiments are the only way of isolating mechanisms to determine their existence and effects (what is called retroduction), but they are impossible in the social sciences. Therefore, they turn their attention to conceptual abstraction and the construction of transcendental arguments (arguments in which we ask what must be true in order for x to exist). Analytical sociologists recognize the important role of experiments in the social sciences for retroductive purposes, but they mainly point to the paramount relevance of agent-based models.
(x) For critical realists, empirical work has a role in the analysis of how different mechanisms interact in a particular situation (what is called *retrodiction*). However, their recommendations in the field of research techniques are limited to the easy appeal to combining intensive and extensive research. While analytical sociologists, in line with the generativist conception of explanation, recognize the potential of computer simulations for retrodictive purposes, critical realists dismiss them, mainly as a result of the epistemological incompatibility between their philosophy and the epistemological implications of agent-based models.

ACKNOWLEDGEMENTS

We would like to thank Daniel Little, Uskali Mäki, Gianluca Manzo, David Miller and Stephen Park Turner for providing valuable remarks on the topic of this article. All errors remain ours.

REFERENCES

Archer, M. (1995), *Realist Social Theory: The Morphogenetic Approach*, Cambridge: Cambridge University Press.
Archer, M. (1998), 'Realism in the Social Sciences', in M. Archer et al. (eds), *Critical Realism. Essential Readings*, London and New York, Routledge.
Archer, M. et al., 'What is critical realism? ASA Theory Section Newsletter, December 23, 2016.' *ASA Theory Section Newsletter*, 23 December 2016. http://www.asatheory.org/current-newsletter-online/what-is-critical-realism.
Antiseri, D. (2006), 'Epistemology and hermeneutics', in M. Alai and G. Tarozzi (eds), *Karl Popper Philosopher of Science*, Rubbettino, Soveria Mannelli.
Bearman, P. and P. Hedström (2011), 'What is Analytical Sociology all about? An introductory essay', in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford University Press.
Bhaskar, R. (1975) [1997], *A Realist Theory of Science*, London: Verso.
Bhaskar, R. (1979), *The Possibility of Naturalism*, Sussex: Harvester.
Bhaskar, R. (1998) [1979], *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*, London and New York: Routledge.
Bhaskar, R. (2009) [1986], *Scientific Realism and Human Emancipation*, London and New York: Routledge.
Bhaskar, R. and A. Callinicos (2015), 'Marxism and critical realism. A debate', *Journal of Critical Realism*, **1** (2), 89–114.
Blom, B. and S. Morén (2011), 'Analysis of generative mechanisms', *Journal of Critical Realism*, **10** (1), 60–79.
Boudon, R. (1971), *Uses of Structuralism*, London: Heinemann.

Boudon, R. (1986), *Theories of Social Change: A Critical Appraisal*, Berkley, CA: University of California Press.
Boudon, R. (1994), *The Art of Self-Persuasion*, Cambridge, UK: Polity Press.
Boudon, R. (1995), *Le juste et le vrai*, Paris: Fayard.
Boudon, R. (2001), *The Origins of Value*, Piscataway, NJ/London: Transaction Publishers.
Boudon, R. (2007), *Essais sur la théorie générale de la rationalité*, Paris: Puf.
Boudon, R. (2009), *La rationalité*, Paris: Puf.
Boudon, R. (2013), *Sociology as Science. An Intellectual Autobiography*, Oxford: The Bardwell Press.
Boudon, R., and F. Bourricaud (eds) (1990), *A Critical Dictionary of Sociology*, Chicago, IL: University of Chicago Press.
Bouvier, A. (2011), 'Individualism, collective agency and the "micro-macro relation"', in I.C. Jarvie and J. Zamora-Bonilla (eds), *The SAGE Handbook of the Philosophy of Social Sciences*, London: SAGE, 199–216.
Bronner, G. and F. Di Iorio (2018), *The Mystery of Rationality*, Springer: Chaim.
Bulle, N. (2019), 'Methodological individualism as anti-reductionism', *Journal of Classical Sociology*, **19** (2), 167–84
Bulle, N. and D. Phan (2017), 'Can analytical sociology do without methodological individualism?', *Philosophy of the Social Sciences*, **47** (6), 379–409.
Chalmers, A. (1988), 'Is Bhaskar's realism realistic?', *Radical Philosophy*, **49** (1), 18–23.
Chen, S-H. (2020), 'On the ontological turn in economics: The promises of agent-based computational economics', *Philosophy of the Social Sciences*, **50** (3), 238–59.
Collier, A. (1994), *Critical Realism: An Introduction to Roy Bhaskar's Philosophy*, London and New York: Verso.
Cruickshank, J. (2007), "The usefulness of fallibilism in post-positivist philosophy: A Popperian critique of critical realism", *Philosophy of the Social Sciences*, **37** (3), 263–88.
Danermark, B., M. Ekstrom and L. Jakobsen (1997) [2005], *Explaining Society: An Introduction to Critical Realism in the Social Sciences*, London and New York: Routledge.
Dawe, A. (1970), 'The two sociologies', *British Journal of Sociology*, **21** (2), 207–18.
Demeulenaere, P. (ed.) (2011), *Analytical Sociology and Social Mechanisms*, Cambridge, UK: Cambridge University Press.
Di Iorio, F. (2015), *Cognitive Autonomy and Methodological Individualism*, Springer: Cham.
Di Iorio, F. (2016), 'World 3 and methodological individualism in Popper's thought', *Philosophy of the Social* Sciences, **46** (4), 352–74.
Di Iorio, F. and S-H. Chen (2019), 'On the connection between agent-based simulation and methodological individualism', *Social Science Information*, **58** (2), 354–76.
Di Nuoscio, E. (2018), *The Logic of Explanation in the Social Sciences*, Oxford: Bardwell Press.
Elder-Vass, D. (2005), 'Emergence and the realist account of cause', *Journal of Critical Realism*, **4** (2), 315–38.
Elder-Vass, D. (2007), 'Social structure and social relations', *Journal for the Theory of Social Behaviour*, **37** (4), 463–77.
Goldthorpe, J.H. (2000), 'Sociological ethnography today: problems and possibilities', J.H. Goldthorpe (ed.), *On Sociology*, Stanford University Press, pp. 65–93.
Goldthorpe, J.H. (2001), 'Causation, statistics, and sociology', *European Sociological Review*, **17** (1), 1–20.
Gorski, P. (2009), 'Social "mechanisms" and comparative-historical sociology: A critical realist proposal', in P. Hedström and B. Wittrock (eds), *Frontiers of sociology* (Vol. 11), Leiden, Boston: Brill Publishers.
Gorski, P. (2013), 'What is critical realism? And why should you care?', *Contemporary Sociology*, **42** (5), 658–70.
Grondin, J. (2002), 'Gadamer's basic understanding of understanding', in R.J. Dostal (ed.), *The Cambridge Companion to Gadamer*, New York: Cambridge University Press.
Harré, H. Rom. (1972), *The Philosophies of Science*, Oxford, GB: Oxford University Press.
Hayek, F.A. (1952) [1948], *The Counter-Revolution of Science: Studies on the Abuse of Reason*, Indianapolis, Liberty Press.
Hedström, P. (2005), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge University Press.
Hedström, P. and L. Udehn (2009), 'Analytical sociology and theories of the middle range', in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford University Press.
Humphreys, P. (2009), 'The philosophical novelty of computer simulation methods', *Synthese*, **169** (3), 615–26.
Kincaid, H. (2017), 'Philosophy without borders, naturally: an interview with Harold Kincaid', *Erasmus Journal for Philosophy and Economics*, **10** (1), 83–100.
King, A. (2004), *The Structure of Social Theory*, London: Routledge.
Lawson, T. (1997), *Economics and Reality*, London and New York: Routledge.
Lawson, T. (2003), *Reorienting Economics*, London and New York: Routledge.
Lawson, T. (2014), 'Popper and social explanation', in T.A. Boylan and P.F. O'Gorman (eds), *Popper and Economic Methodology*, London and New York: Routledge.
León-Medina, Francisco J. (2014), 'Reasons and biological causes. Some reflections on Boudon's Theory of Ordinary Rationality', *Papers. Revista de Sociologia*, **99** (4), 595–629.
León-Medina, Francisco J. (2017), 'Analytical sociology and its 'syntax' for explanation', *Social Science Information*, **56** (4), 502–25.

Linares, F. (2018), 'Agent based models and the science of unintended consequences of social action', *Revista Española de Investigaciones Sociológicas*, **162**, 21–38.
Little, D. (2013a), 'Bhaskar's core ideas', *Understanding Society*, Tuesday, 3 December 2013, https://understanding-society.blogspot.com/2013/12/bhaskar-core-ideas.html
Little, D. (2013b), 'Cruickshank's central critique', *Understanding Society*, Thursday, 5 December 2013, https://understandingsociety.blogspot.com/2013/12/cruickshank-central-critique.html
Little, D. (2013c), 'Replies by Elder-Vass and Hartwig to Cruickshank/Little', *Understanding Society*, Thursday, 12 December 2013, https://understandingsociety.blogspot.com/2013/12/replies-by-elder-vass-and-hartwig-to.html
Little, D. (2013d), 'Cruickshank on philosophical issues with critical realism', *Understanding Society*, Wednesday, 13 November 2013, https://understandingsociety.blogspot.com/2013/11/cruickshank-on-philosophical-issues.html
Little, D. (2013e), 'Kaidesoja's naturalistic social ontology', *Understanding Society*, Sunday, 15 December 2013, https://understandingsociety.blogspot.com/2013/12/kaidesojas-naturalistic-social-ontology.html
Little, D. (2016), 'What is critical realism?', *Perspectives. A Newsletter of the ASA Theory Section*, 23 December 2016.
Little, D. (2018), 'What is social ontology and why do we need it?' Interview with Daniel Little by Tim Rutzou in *Context of the Critical Realism Project*, directed by Phil Gorski', unpublished paper.
Little, D. (2020), 'Social Ontology De-dramatized', *Philosophy of the Social Sciences*, https://doi.org/10.1177/0048393120916145.
Lizón, Á. (2016), *La sociabilidad humana: estudios de evolución y cognición social*, Síntesis.
Lukes, S. (1968), 'Methodological individualism reconsidered', *The British Journal of Sociology*, **19** (2), 119–29.
Mäki, U. (2011), 'Scientific realism as a challenge to economics (and vice versa)', *Journal of Economic Methodology*, **18** (1), 1–12.
Mäki, U. and P. Oinas (2004), 'The narrow notion of realism in human geography', *Environment and Planning A*, **36** (10), 1755–76.
Manicas, P.T. (2006), *A Realist Philosophy of Social Science: Explanation and Understanding*, New York: Cambridge University Press.
Manzo, G. (2007), 'Comment on Andrew Abbott/2', *Sociologica*, **1** (2), 1–8.
Manzo, G. (2010), 'Analytical sociology and its critics', *European Journal of Sociology/Archives européennes de sociologie*, **51** (1), 129–70.
Manzo, G. (2013a), 'Is rational choice theory still a rational choice of theory? A response to Opp', *Social Science Information*, **52** (3), 361–82.
Manzo, G. (2013b), 'Educational choices and social interactions: A formal model and a computational test', *Comparative Social Research*, **30**, 47–100.
Manzo, G. (2014), 'Data, generative models, and mechanisms: more on the principles of analytical sociology', in Manzo, G. (dir.), *Analytical Sociology: Actions and Networks*, Chichester: John Wiley.
Manzo, G. (2020), 'Agent-based models and methodological individualism: Are they fundamentally linked?', *L'Annee sociologique*, **70** (1), 197–229.
Miller, K.D. (2015), 'Agent-based modeling and organization studies: A critical realist perspective', *Organization Studies*, **36** (2), 175–96.
Noguera, J.A. (2003), '¿Quién teme al individualismo metodológico?: Un análisis de sus implicaciones para la teoría social', *Papers: revista de sociologia*, **69**, 101–32.
Popper, K.R. (1957), *The Poverty of Historicism*, Boston, MA: Beacon Press.
Popper, K.R. (1963), *Conjectures and Refutations: The Growth of Scientific Knowledge*, London: Routledge and Kegan Paul.
Popper, K.R. (1980),'Autointerpretazione filosofica e polemica contro i dialettici', in C. Grossner (ed.), *I filosofitedeschi contemporanei* (Verfall der philosophie. Politik deutscher Philosophen), Rome: Città nuova.
Sayer, R.A. (1984) [1992], *Method in Social Science: A Realist Approach*, London: Psychology Press.
Törnberg, A. (2019), 'Abstractions on steroids: A critical realist approach to computer simulations', *Journal for the Theory of Social Behaviour*, **49** (1), 127–43.
Turner, S.P. (2013), 'The Philosophy of the Social Sciences in Organizational Studies' in S.R. Clegg, C. Hardy, T.B. Lawrence and W.R. Nord (eds), *The SAGE Handbook of Organization Studies*, Newbury Park: SAGE, pp. 409–24.
Weber, M. (1978) [1922], *Economy and Society*, Berkeley, University of California Press.

Francesco Di Iorio and Franceso J. León-Medina - 9781789906851

7. Analytical sociology and cultural sociology[1]

Filippo Barbera

1. THE WAR OF THE ROSES

The relationship between analytical sociology and cultural sociology is not, *prima facie*, a fruitful one. The *esprit de géométrie* that characterizes the former would appear not to agree with the *esprit de finesse* that marks the latter. While analytical sociology underlines the role of the mechanisms-based explanation, the beacon of cultural sociology is the principle by which social action should be interpreted as a text, and the interpretation depends on the network of meanings that constitute the context of the action. The gap between explanatory science and hermeneutical disciplines is here the deepest. To cut through a vast theoretical underbrush, it could be argued that the aim of cultural sociology is the construction of *thick descriptions* that are able to seize the cultural meaning of actions to spell out their wide-ranging implications, as well as to highlight the role played by the constitutive rules of social interaction. Analytical sociology, on the other hand, favours *thin descriptions* (Hedström and Udéhn 2009) and discusses the problem of meaning in an ancillary way as *organon* of the explanation. The interpretation of meaning becomes either a specific phase of the research process (Goldthorpe 2000) or a component of a micro-founded theory focused on the cognitive complexity of decision-making processes (Esser 1993; Lindenberg 1993, 2000). These wide-ranging and "foundational" differences become less divisive in those empirically oriented pieces of works that have both explanatory and interpretative aims. On the part of cultural sociology, for example, this is the case of the so-called endogenous relevance of the cultural processes (Kaufman 2004) and of the analysis focused on how cultural tastes shape social networks (Lizardo 2006). On the part of analytical sociology, empirical works such as *Chain of Affection* by Bearman et al. (2004) do the same job and help to close the gap. True as it may be, these attempts do not speak for the mainstream trends, and the dialogue between analytical sociology and cultural sociology is still quite a challenge for social theory. As will be illustrated, to bridge the cliff it is key to properly portray the merits and pitfalls of contenders, while avoiding getting caught in endless ontological debates about the "true nature" of social phenomena. For analytical sociology, paying attention to the sirens of ontology ("Is *the real* relational"?) can doubtless be useful, but only if binding oneself to the theoretical and methodological consequences of this standpoint, *never* as a self-sufficient question. As any voyage on the seas of contemporary sociology can testimony, explorers less cautious than Ulysses have left many a battered wreck behind.

The first divide is determinism vs. intentionality. At first glance, cultural sociology would appear to support a deterministic framework where the social actor is pushed by forces that drive them toward "choices" whose main outcome is the reproduction of the social system. This standpoint denoted both the Parsons' version and Bourdieu's account of the "cultural effects",

[1] I am grateful to Krzysztof Krakowski for his comments to an earlier version of the chapter.

Filippo Barbera - 9781789906851

albeit with quite opposing aims. Parson's account results in the idea of the social actor as a *cultural dope* who follows the social norms and the values embedded in his personality by the socialization process (Wrong 1961). In the oversocialized conception of "man", individuals are presumed to acquire culture in the course of socialization and to enact it unproblematically (DiMaggio 1997). The interiorization of norms and values has a positive effect on the reproduction of the social system, since it allows social action to follow the role expectations in the correct way (Parsons 1937). Bourdieu's version, on the contrary, expresses the idea that culture induces the actor to act *against* its interest, thanks to the effects resulting in the positioning in a field that influences its dispositions, *habitus* and cultural capital (Bourdieu 1984). The reproduction of the system is hence tightly connected to the reproduction of power and inequalities. Is this a useful starting point to promote a constructive dialogue between analytical and cultural sociology? Not really. It is fair to say that today almost nobody—in the field of cultural sociology—would support a naïve Parsonian functionalism. Additionally, the deterministic interpretations of Bourdieu appear to be the fruit of a specific intellectual climate, heavily marked by Marxism and by the prevalence of class analysis, more than the faithful mirror of the faceted thought of the French sociologist (DiMaggio 2007). As Paul DiMaggio (1997, p. 265) clearly stated in connection to Ann Swidler's (1986) definition of "culture as toolkit": "The view of culture as values that suffuse other aspects of belief, intention, and collective life has succumbed to one of culture as complex rule-like structures. However, it is still true that a number of contemporary approaches focused on the "cultural effects" tend to emphasize the role of the symbolic dimensions as *forces* that shape the course of the action *pushing* individual volition and cognition. This is the case, for example, of the so-called "ideational" dimension of the power of public discourses and its influence on the cognitive and normative maps of the actors. Such an influence would unravel with the power *through ideas*, namely in connection to the capacity of the actors to convince other actors to accept and adopt their opinions on what to think and do; through the power *on ideas*, tied to the imposition of specific ideas in the decision-making arenas, and through the power *in ideas*, which concerns the institutional power processes and the sub-cognitive and pre-reflexive functioning mechanisms of the institutions (Carstensen and Schmidt 2016). The happy wedding between the sociology of cultural processes and the tradition of critical analysis of power—from Foucault to Gramsci—is a key component of these approaches. A happy wedding, this one, that tends to dismiss the interpretive space and the transformative agency that—pace "biopolitics" and "hegemony"—open the door to a pattern of interactions that pragmatically *decouple* individual meanings and actions from macro-level symbolic and discursive "logics". For instance, the work of David Stark and collaborators, creative friction in folded structures is a chief case in point (2009; 2017). Folds are cliques that are joined by sharing a single node (de Vaan et al., 2015; Vedres and Stark, 2010) and, from the analytic standpoint, the greater the *symbolic* distance between the cliques, the more their partial overlapping through structural folds will be a potential source of *transformative agency* (Sewell 1996, 2005), bringing diverse elements together. The so-called cognitive turn in cultural sociology (Di Maggio 1997) and the political and moral sociology of worth (Boltanski and Thévenot 2006; Boltanski and Vitale 2006) point exactly toward this direction and are closer, at least in principle, to the analytical sociology principles.

Analytical sociology, for its part, is rightly portrayed as having the propensity for a theory of action and interaction which looks at the role of *mental states* and models the *social structure* in which the interaction unfold (Lizardo 2012). This portrait is correct (Manzo 2014). What is both unfair and not very useful for a constructive dialogue is to deny that an intentionalism

Filippo Barbera - 9781789906851

of this kind is different from rational choice theory, a point that analytical sociologists have forcefully argued with convincing arguments (see Manzo 2010 and 2013). The intentionalism of analytical sociology stands out from rational choice theory, which justifies the adoption of non-realistic assumptions, either on the basis of the predictive accuracy or in connection to the analytical power and parsimony. The micro-level reference point is the so-called DBO "theory," which states that the proximate cause for action is a particular combination of desires, beliefs and opportunities (Hedström 2005). What needs to be underlined is that the analytical status of the DBO "theory" is far from clear, standing in between a general framework and a propositional theory. The difference between frameworks and theories states that a framework helps to identify the elements and relationships among these elements that one needs to consider, while theories enable the analyst to specify which elements of the framework are particularly relevant to certain kinds of questions and to make general working assumptions about these elements (Ostrom 2005). Several theories are thus *compatible* with a single framework: for instance, in principle the DBO framework is compatible with the rational-choice theory, the interaction-ritual chain theory, with cognitive dissonance reduction, with learning theories and so on.

What should be chiefly emphasized in this connection is that—in one way or another—a micro-level framework marks the distance with the *macro-narratives* toward which cultural sociology tends to indulge. Producing lists of cunning conjectures does not imply sound explanatory statements: an inventory of plausible factors can easily turn into a set of *well-written* narratives where everything is equally key for the explanation of the phenomena. In other words, macro-level narratives do not specify the *how mechanisms* through which the macro-level change (or stability) occurs. Producing a long list of plausible cultural factors, dimensions and variables such as group values, roles, prestige patterns, power relations, emotional dynamics, rituals and organizational cultures is certainly useful for the logic of invention and to stimulate new research ideas (Abbott 2004), but it is hardly satisfactory from an explanatory viewpoint. It follows that the plausible processes invoked should make sense of the fine-grained mechanisms or "powerful particulars"—to borrow a critical realist notion in line with the analytical sociology perspective—able to generate the phenomena of interest. Every causally relevant research program must include a syntactical dimension and not just a semantic one (Abell 1987). The painless elaboration of factors and dimensions of various kinds lacking a syntactical backbone would obscure—rather than clarify—social processes. For instance, it is not enough to argue that the nexus between the Protestant ethic and the spirit of capitalism must "contextualize" the Reformation in broader *socio-cultural* processes, such as (i) the weakening of old roles and the emergence of new ones, less constrained by "liturgical" routines and codes; (ii) the decline of imperial legitimacy and the rise of nation-states; (iii) the strengthening of secular normative orders in place of religious ones; and, eventually, (iv) the new mundane obligations correlated to the emergent set of duties and salvations do not help much in reaching causal depth (Jepperson and Meyer 2011; see the critical remarks of Abell, Felin and Foss 2014 and the comment of Barbera and Negri 2014).

This slant of cultural sociology would appear to be supported by a further happy wedding,[2] namely the idea that history and sociology are and always have been the same thing (Abrams 1983), with the consequence that its practitioners "enjoy a delightful freedom to play 'pick

[2] It follows that cultural sociology might be accused of bigamy for its double happy wedding. In this regard, I feel comfortable in proposing a third happy wedding.

Filippo Barbera - 9781789906851

and mix' in history's sweet shop" (Goldthorpe 2000, p. 43). From the analytical sociology standpoint, the most important critique to this posture is that the effortless construction of a list of potentially important dimensions is a *waste of time*. What matters is the outline of empirically testable generative mechanisms, which are sufficient to generate a macro-level phenomenon of interest (Epstein 2006). For analytical sociology, an "external reality" exists and it is precisely in the tension that exists between the unattainability of complete knowledge of this reality and a liking for knowledge that is effective even if *incomplete* that we can place the interest in sociological knowledge. Reality can only be grasped in a selective manner. A complete cartography of the social world is both ridiculous and useless, as J.L. Borges aptly argued (1998) in the "paradox of a perfect cartography". Even sheer description always entails selection based on a relevance criterion, namely selecting details that are most appropriate to the topic at hand (Sen 1980). Excessive details and endless lists of potentially key dimensions are virtually paralyzing for understanding society, since they hinder abstraction and reduce a sociologist to a mnemonist who lived in a world of unique—but sadly *powerless*—particulars (Barbera 2011).

2. SLEEPING WITH THE ENEMY

I have been arguing that the differences between analytical sociology and cultural sociology are certainly important, but they do not represent insurmountable barriers. I have been illustrating what these divides are about and what should be underlined for a constructive dialogue to start. To really bridge the gulf, however, a certain amount of *intellectual risk* is necessary and—as we will see—the premium for the risk is higher when shifting from the level of "foundational" principles to the one of theoretically guided empirical research. In this respect, I will illustrate three research papers that take into account cultural phenomena and processes. These three papers are selected from the list of winners of the "Robert K. Merton Award", the annual award given by International Network of Analytical Sociologist (INAS) to the best paper in the spirit of analytical sociology and middle-range explanation (for a comprehensive analysis of this award, see Manzo, this volume). We can consider these three papers as sort of *critical case studies* for the relationship between analytical sociology of cultural sociology. Since the papers were awarded by the community of analytical sociologists, they represent the quality conventions of analytical sociology very closely and, if the gulf can be bridged here, it is likely to be bridged anywhere (Table 7.1).

The three winners that are more closely focused on cultural sociology themes are the winners of the 2017, 2018 and 2019 editions.

2.1 Deviance as Situational Choice

The paper entitled "Disorder, social capital, and norm violation: Three field experiments on the broken windows thesis" by Marc Keuschnigg and Tobias Wolbring (2015) relies on a weak methodological individualism perspective (Udéhn 2001) to explain how and why the contextual *signals*[3] may have a deviant effect on actors' behaviours and decisions.

[3] For the sake of precision, these would be better called "signs" and not "signals". The latter are intentionally emitted for strategic purposes, as in the Ozan Aksoy and Diego Gambetta's paper that I will illustrate.

Filippo Barbera - 9781789906851

Table 7.1 Papers that were winners of the R.K. Merton Award 2013–2019

	2019	2018	2017	2016	2015	2014	2013
Papers	Amir Goldberg and Sarah K. Stein, "Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation". *American Sociological Review*, 2018. Vol. 83, Issue 5, pp. 893–932.	Ozan Aksoy and Diego Gambetta, "Behind the Veil: The Strategic Use of Religious Garb". *European Sociological Review*, Volume 32, Issue 6, December 2016, pp. 792–806.	Marc Keuschnigg and Tobias Wolbring, "Disorder, Social Capital, and Norm Violation: Three Field Experiments on the Broken Windows Thesis". *Rationality and Society* 2015. Vol 27, Issue 1, pp. 96–126.	Sarah K. Cowan, "Secrets and Misperceptions: The Creation of Self-Fulfilling Illusions". *Sociological Science*, 2014, Vol. 1, pp. 466–92.	Elizabeth Eve Bruch, "How Population Structure Shapes Neighborhood Segregation." *American Journal of Sociology*, 2014, Vol. 119, Issue 5, pp. 1221–78.	Javier Polavieja, "Socially-Embedded Investments: Explaining Gender Differences in Job-Specific Skills". *American Journal of Sociology*, 2012, Vol. 118, Issue 3, pp. 592–634.	Paul DiMaggio and Filiz Garip, "How Network Externalities Can Exacerbate Intergroup Inequality". *American Journal of Sociology*, 2011, Vol. 116, Issue 6, pp. 1887–933.

Filippo Barbera - 9781789906851

The research is an application of the so-called "broken window" theory, a well-known topic in the sociology of deviance that traces the norm-breaking behaviour back to ongoing situational features instead of relying on the role of pre-existing individual attitudes: "(the theory) states that physical and social disarray fosters additional norm violations and, therefore, the spreading of disorder […] Put into practice, (the theory) advocates the removal of signs of disorderliness and incivility in order to curb crime. Empirical findings, however, were rather ambiguous […] and did not allow decisive conclusions about causality" […] (p. 97). In the spirit of analytical sociology, the authors are interested in the explanation of the phenomena (deviance) through the rigorous empirical control of micro-founded hypotheses. For this purpose, they set up a natural experiment research design aimed at measuring the behaviours in ordinary day-to-day situations. The micro-level hypothesis constitutes a variation of Gary Becker's *rational-choice* model (1976), adapted to situations where the expected costs associated to social norm violation are low, for example, throwing a small piece of scrap on the ground instead of the bin, crossing the road against a red light and keeping an envelope/letter found on the ground with a variable amount of money. In these situations, where deviance is *low cost*, social actors derive their expectations/beliefs on the probability of being punished from contextual evidence, hence again from "signals" resulting from the macro-micro situational context of the action (see Hedström and Swedberg 1998). Those signals have proximal characteristics, for example, the level of cleanliness of the environment in which the action takes place and distal characteristics, like the social capital level of the reference setting (district, community etc.). The choice of violating the norm is, in low-cost situations, mediated by heuristics used to "interpret" the situational signals and not through a cost-effective calculation. In *low-cost* situations then, "the individual decision maker has no direct interest in the outcome because she or he will hardly be influenced by the outcome. Hence, when individual costs and benefits are low it is rational to shorten the lengthy and cumbersome decision-making process by the use of heuristics" (Keuschnigg and Wolbring, 2015, 97). As for Baudelaire in "The Flowers of Evil", "In Nature's temple, living columns rise, which oftentimes give tongue to words subdued, and Man traverses this symbolic wood, which looks at him with half familiar eyes."

In the first field experiment, the authors measure of an environment characterized by the presence of waste increases the penchant of individuals not to follow the mundane norms that prescribe "not throwing scrap paper on the ground". While leaving school or home, we find a useless leaflet left on our bicycle, parked in a car park with a waste bin. In one case, the context of the action is clean and there is no trace of "disorder" and carelessness; in the second case, on the contrary, we are surrounded by scrap paper and a poorly cared for environment. Furthermore, the two contexts also differ for the social capital level of the neighbourhood: high in one case and moderate in the other. In the second field experiment, the low-cost deviant behaviour refers to crossing the road against a red light, stimulated by the eyesight of others who violate the norm in crowded and potentially open places characterized by different levels of social capital. We are in a peripheral neighbourhood, with little cared for surroundings, low electoral participation, few associations and little social cohesion. We are at a traffic light, waiting for the green light, close to a crowded place like a bus stop. A person comes up beside us and does not wait for the green light to cross the street. Do we decide to do the same? Let us repeat the experiment in a central neighbourhood, in a clean environment and with high social capital. What changes in our

Filippo Barbera - 9781789906851

behaviour? The third experiment re-proposes the well-known case of the lost letter (Milgram et al., 1965; Simmons and Zumpf, 1983), in relation to the presence of a systematic variation of the degree of situational disorder and social capital. We see an envelope on the ground, next to a letterbox, with a window showing a 10 or 100 euro banknote. Let us reproduce the situation in two urban contexts, different for social capital and environmental disorder, and measure the behavioural variations accordingly. Briefly, the three field experiments results are consistent with the micro hypotheses spelled out: a situation characterized by disorder and little care induces individuals to throw the leaflet to the ground, to cross the street against the red light and to keep the letter found on the ground. The distal social capital, moreover, has a protective effect: the violations of social norms took place less frequently in areas with high endowments of social capital. Additionally, in the third experiment, and in line with the micro-level hypotheses, no effect of "disorder" was noted on the decision to steal 100 euro banknotes. This suggests that the individuals worry *less* about contextual signals in situations in which the expected cost of the punishment is high. With the increase in expected costs and benefits, the "symbolic wood" and "living columns" of Baudelaire's poem become less important for individual choices.

2.2 The Veil as Signal and Self-Binding

The second award-winning paper we consider was written by Diego Gambetta and Ozan Aksoy (2016) titled "Behind the Veil: The Strategic Use of Religious Garb". The paper deals with a classic problem of cultural sociology, which is the relationship between modernization and attitudes/behaviours consistent with religious doctrine. The focus of the analysis are the various head covers and dresses worn by Muslim women and collectively referred to as "veiling". Different theoretical propositions would outline a tendency toward veiling less, as well as a more marked secularization of behaviours connected to higher education, income and social class. Hence, the first hypothesis of the research—in line with the theory of cultural modernization and secularization—identifies a negative relationship between exposure to the values and dimensions of secularization (education, income and urbanization) and veiling. This "classic" hypothesis, however, clashes with the presence of puzzling phenomena: "Yet, much anecdotal evidence indicates that from the late 20th century onward veiling spread chiefly among the highly educated, young, urban, middle-class women" (p. 792). To account for these puzzling phenomena, the authors rely on precise micro-founded hypotheses: the decision to wear a veil (and/or a more or less favourable attitude toward veiling) is a strategic *signal* that women decide to "convey, either with the intent to manage their behaviours or to communicate their piousness to their communities. Veiling, as such, should be framed as a response to the 'temptations' posed by social circumstances. With modernization, women are increasingly exposed to contacts outside of their families and employment, and this could put women, especially younger, urban, and single women, at a greater risk of abandoning their traditional mores and restraint, or even just to be believed to be doing so" (p. 793). Veiling as a strategy could follow two paths: first, it could be used as a self-binding device against temptation (Elster, 1979). In this case, it would work as a safeguard of women's modesty outside the home and a deterrent against ill-intentioned men's approaches. Second, it could be directed toward others as a strategy to communicate one's persisting piety in the face of modernization's challenges (p. 793). Hence veiling is both a self-binding commitment and a

Filippo Barbera - 9781789906851

strategic signal whose probability is higher among religious women who interact in risky and high-temptation environments.[4] Additionally, the context influences veiling in connection to the "neighbourhood" behaviour of reference groups. In particular, as the third hypothesis maintains, the spreading of the veil is a signal that women use to hint at their knowledge of social norms and the probability of being sanctioned in case of non-compliance, so "where the average veiling is higher and veiling variation is lower, we can expect that veiling will be higher" (p. 795). The effect of the context has two implications, which give rise to additional hypotheses. One is that in the Western contexts veiling will be less spread the higher the presence of indigenous women in the neighbourhood and the higher the number of friendships with native people. Likewise, context-related characteristics constitute "temptations" and risks to which the more educated women respond with self-binding/strategic external communication strategies, which result in a higher probability of veiling. The empirical test of these hypotheses resorts to three quantitative data sets, two of which include information on veiling among Turkish women in two countries: Belgium, where Muslims are a minority, and Turkey, a relatively secularized country but with a Muslim majority in the population. The third data set is made up of a representative sample related to 25 Muslim countries, among which the authors select those countries which have experimented more marked secularization processes, where women were not only exposed to religious customs and values and where the veil also represents an intentional choice, not just a cultural obligation. The three data sets return different measurements of the phenomena of interest: "The Turkish survey […] asks not only whether a woman veils outside the home but also in which of four forms of increasing strictness: no veil, headscarf, turban and chador. The Belgium dataset measures attitudes towards veiling with the question: Muslim women should cover their head when outside home' (one=completely disagree to six=completely agree). In the Pew dataset, the interviewer records whether the respondent was veiled during the interview, (using four categories: no veil, hijab (similar to turban), niqab (similar to chador), and burqa (2016, 796)". Resorting to quantitative analysis techniques, the authors find a first empirical corroboration to the illustrated hypotheses. In particular, the first hypothesis is in agreement with the results obtained in all three data sets, confirming the "classic" theory of cultural modernization. The analysis results, furthermore, are in line with the second hypothesis, both in Turkey and in Muslim countries, but not in the case of the Belgium data set. Additionally, empirical findings are in line with the third and fourth hypothesis: the average veiling in one area is positively associated with the propensity to veil for the averagely religious women, while highly religious women seem much less responsive to the prevalence of veiling as this makes the environment less threatening: "[…] greater variation in veiling has either a negative or non-existent effect for the averagely religious women […], while for the highly religious women it increases veiling […] (2016, p. 800). As a general trend, Aksoy and Gambetta conclude that veiling decreases as modernization increases; but to account for the veiling behaviour of the highly religious "the commitment and signalling theories are needed, as veiling does not seem to be a mere expression of religiosity but the result of women's strategic decision" (2016, 802).

[4] As the authors maintain, this hypothesis is conditional on high religiosity, thus it does not refer to the main effects (in the statistical sense) of the variables, but to their interactions with religiosity.

Filippo Barbera - 9781789906851

2.3 Cultural Differentiation and the Meaning of the Action

The third article, winner of the R.K. Merton award for 2019, is "Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation", by Amir Goldberg and Sarah K. Stein. The starting point of the paper is the systematic pervasiveness of cultural differentiation: musical tastes, political preferences, religious beliefs and ideas of justice form clusters marking borders and "worlds", building correlated patterns to social position, status and various types of inequalities. How do these "cultural worlds" form? The answer of sociologists has often concerned *social contagion* models, epidemiological models whose cultural differentiation comes either from the previous presence of segregated social structures or from the tendency of actors to interact among peers according to the rules of homophily and preferential attachment.

Structural differentiation—spatially segregated groups or high-density networks separated from other networks—tends to favour imitation, conformism and adoption of accepted practices among "peers" and "neighbourhood". The presence of cross-cutting ties *among* groups, alone, is not enough to favour merging between different meanings and practices, as it typically happens when *multiple* concomitant sources of influence are necessary (see the case of the complex contagion, Centola and Macy 2007; Centola 2018). The second mechanism, on the other hand, connects the cultural differentiation with the trend of interacting with peers, for both intrinsic and instrumental reasons. In this way, however: "Cultural differentiation […] is ultimately epiphenomenal to a structurally segmented world" (2018, p. 898).

The social contagion epidemiological models do not distinguish between the adoption of a practice (e.g. getting a vaccine) and its interpretation (what getting a vaccine *means*). In order to focus on the generative mechanisms connected to the interpretation and, lastly, to the meaning of cultural practices, the paper adopts an approach which analyses culture starting from the functioning of cognitive processes: "meaning inheres in cognitive associations between concepts and perceived associations constrain people's actions […] People learn from their social environments how to associate between different cultural practices, and in their own behaviors enact, and therefore reproduce for others, these associations" (2018, 898). Additionally, Goldberg and Stein decouple the genesis and spreading of the meaning of cultural practices from the presence/absence of interpersonal ties. The idea is that social ties and networks matter "when the cultural knowledge being shared is complex or costly" (2018, 901), but that is not always the case. Like in the symbolic interactionism and phenomenology perspectives (Berger and Luckmann 1967), communication is chiefly understood as a carrier of meaning even just in the presence of superficial interactions and without calling into question form and/or content of the social ties and networks. To make this key point clear, the article mentions the results of a field experiment by Paluck and Shepherd (2012), designed to estimate the effects of peers on high-school bullying. If the adoption of potentially *high-cost* behaviours, such as defending bullied classmates, is mediated by strong ties and interpersonal relations, the mere exposure to the lesser impact of bullying is enough to reduce the probability of behaving like bullies, notwithstanding the fact that the peers under review were friends or just classmates.

If the "carrier" of the spread does not depend on social networks and the type/quality of ties, how do you generate cultural differentiation? To answer this question, it is necessary to leave the epidemiological analogy, which considers cultural transmission as a *content-free* broadcast-receiving mechanism mediated by "channels". The cultural transmission/

Filippo Barbera - 9781789906851

differentiation requires primarily understanding the interpretation process through which the individuals put the practices/actions in relation with their semantic content. For example, "Cigarettes, cigars, and pipes provide similar physiological utilities, but these different forms of tobacco consumption are commonly associated with distinct cultural meanings. Whereas cigars connote masculinity and power, pipes are conventionally associated with contemplation and old age" (2018, p. 903). In this way, the authors distinguish between the behaviour (e.g. smoking) and its cultural meaning (e.g., smoking is an act of rebellion). The cultural diffusion is articulated as a consequence in a two-stage process accordingly: the actor *interprets* other people's behaviour and then *evaluates* such behaviour. The interpretation is functional to the association of concepts, whose meaning derives from their relations of implication and opposition. The association of concepts comes from the observation of other people's behaviour and from the co-occurrences that may be observed in other people's cultural practices. For example, "A pupil notes that the lads smoking at the school gates are also wearing platform shoes, engaging in physical violence, and generally 'having a laff' at the ear'ole's academic aspirations. Together, these behaviors form a gestalt connoting resistance to the school's establishment and the middle-class ideals it represents" (2018, 906).

The second stage—behaviour evaluation—is based on the *constraint satisfaction*, which considers a connectionist reading of the psychological mechanism of cognitive dissonance reduction. We do not just tend to make our mental states consistent (what I wish to do and what I perceive I can do, like in the fox and the grapes fable), but we also need that meaning relations and the resulting story telling are consistent. The *associative diffusion* is therefore a two-stage process: in the first stage, the actors draw a meaning matrix that connects social practices, observing their co-occurrences in other people's behaviour; in the second stage, they adapt their preferences and behaviours to these co-occurrences, eliminating incongruent information and reducing cognitive dissonance. Additionally, by adopting "perceptively congruent" visible behaviours, the actor contributes to the reproduction of the cultural order. As mentioned before, this perspective does not put at the centre the "social contagion" of behaviours, but the interpretative learning of the *meaning* of social practices. To test this proposition, Goldberg and Stein (2018) implement an agent-based model (ABM). As highlighted by analytical sociology (Manzo 2014), resorting to agent-based models allows us to rebuild *in silico* the generative micro-mechanisms of the macro-phenomenon of interest (the cultural differentiation). The model shows, in line with the outlined theoretical setting, that the agents converge toward an interpretative consent, but they are characterized by different preferences and behaviours.

3. GUESS WHO IS COMING TO DINNER

The three pieces of research just illustrated can be ordered along a *thin description–thick description* gradient. Along this gradient, three key issues to be addressed for the dialogue between analytical sociology and cultural sociology emerge. To begin with, the Keuschnigg and Wolbring paper portrays deviance as an intentional decision affected by the proximal signals of the context and by the distal resources of social capital: the low-cost *framing* triggers choice heuristics resulting into the decision to act against the norms in relation to the context signals or, better, *signs*. The process, as Raymond Boudon (1989, 1993) maintained, is clearly based on good situational reasons: it is reasonable even if it does not follow a costs-expected benefits logic. The action theory used is certainly relevant for cultural sociology,

Filippo Barbera - 9781789906851

since it articulates the action process in an attention/cognition theory and in a decision theory, illustrating how the forming processes of normative beliefs of situational nature are based on "good reasons". Such an approach constitutes, in conclusion, a valid sociological theory of attention or salience, to be integrated with intentional choice models. Intentionality, it should be stressed, requires information, and this is not acquirable without postulating some kind of *cognitive burden* of the agent, either resulting from the cost/benefit computational capacity, or from the learning process grounded on the classification and storing of procedures for problem-solving, or from the ability to compare the effectiveness of one's decision with other people's decisions. Burdensome cognitive capacities are not requested only for those models where actor behaviour does not require information, that is, for those models which consider the action as completely determined by social structures or by internalized cultural models.

Secondly, the Aksoy and Gambetta research concerns religious cultural practices or cultural practices used as self-binding strategies or strategic *signals*, which are triggered because of contextual dimensions that are able to create a set of opportunities and "temptations" for people. Here, too, actors are intentional, have set objectives and act reasonably given the contexts in which they are embedded. Symbolic communication can be the consequence of intentional strategies aimed at signalling the belonging and loyalty to the reference group. The actors can be rationally bound to a course of action, if this is the only way to accomplish their desires and preferences (Elster 1979). Furthermore, it is hard to signal truthfully, hence believably and effectively, one's desire of being a friend (or an espouse) in an insincere way. Strategies, after all, should deal with the practical tension between intentionality and expressive action: many individual states of emotional type may be secondary or unintentional products of other actions. In addition, they are effective for they are expressive and non-goal directed. For example, spontaneity is a state that can be characterized as an "essentially secondary" effect; in fact, it is difficult to be able to be *intentionally* spontaneous. What is exchanged in "social life" is a particular category of goods (Bianchi et al. 2018): symbolic goods that are exchanged between two or more actors under conditions of incompletely specified obligation (Blau 1963). Micro-level social research showed that both group identity and emotional enforcement display a positive effect on commitment and support the exchange of such goods (Lawler and Yoon 1998; Lawler et al. 2000, 2008; Lawler 2001). Relying on previous experimental evidence, Lawler (2001) proposed a causal model to account for the mechanisms that cause the emergence of "a recurrent pattern of interaction among a set of actors, from which they come to *perceive themselves as a unit* (i.e. a group) and to develop feelings about that unit" (Lawler 2001, pp. 4–5, emphasis added). In this line, as Pizzorno (2006) argued, group identity, emotions and expressive behaviour have the task of supporting the system of meanings in which social actors are embedded and the expectations express their self-representation as *persons*. Social goods such as social approval, status, respect, recognition, love, friendship and trust often reward behaviours whose value depends on the *intrinsic meaning* of the action course, rather than on its extrinsic result.

Third, the research by Goldberg and Stein models the social construction of the meaning, drawing from both psychological and socio-anthropological literature. Here, the actors are both *meaning-makers* looking for associative meanings among practices, and *story-tellers* moved by the search for semantic coherence. This perspective is in line with the need to introduce *cultural mechanisms* in the macro-micro-macro schemas (Barbera and Negri 2015), as those approaches deal rigorously with the cultural meaning of actions (Lizardo 2006; DiMaggio 1997). The meaning of social action can and should be included in the macro-micro-macro

Filippo Barbera - 9781789906851

framework, even with regard to its more complex implications such as social identity and non-intentional processes (Pizzorno 2006). This needs a certain degree of intellectual risk on the side of analytical sociology. In this connection, as we will argue in the concluding section, the bottom left-hand corner of the Coleman boat usually implies *pre-formed agents*, without accounting for the cultural and symbolic dimensions of social processes that constitute them *as such*. Instead, agents should be allowed to *emerge* from meaning-making processes, as in the perspective of symbolic interactionism and phenomenology. In this case, intentionality plays a residual role.

4. AFTER THE WEDDING

A wedding is a two-person game. However, even without getting married, a well-played game requires the active cooperation of the parties. Dancing or rowing, for example, require coordination to the other's pace. Without coordination, there is no pleasure nor benefit. The three outlined researches are, in this perspective, a good starting point: the R.K. Merton award, in the last three years, has been given to potentially sympathetic works with a perspective of cultural sociology. That of Goldberg and Stein, in particular, adopts a very close perspective to that of Pierre Bourdieu—often framed in sharp contrast to analytical sociology (Van de Berg 1998; Hedström 2005)—hybridized with micro-founded models focused on *meaning-making* mechanisms. The meaning of social action and its cognate concepts are not alien to an analytical sociology framework; rather they are sometimes boiled down to other concepts, such as cognitive rationality (Boudon 1998), framing (Lindenberg 2000; Esser 1993), signals (Gambetta 2009), beliefs (Rydgren 2009) and still others (Manzo 2010; 2012). From the viewpoint of analytical sociology, the *Maginot* line is—as usual—to keep a safe distance from those explanations that "by decree" waive the role of actors and interactions. In this line, is the analytical sociology syntactical framework—the so-called "Coleman boat/bathtub"—a proper tool to make sense of meaning-making mechanisms? May we use it to address the cultural sociology agenda? Oriel Sullivan et al. (2018) give us a tentative "yes." Their paper deals with the slow change in the division of unpaid work and care (lagged generational change) and the authors argue that "Large-scale social changes involve complex patterns of relations between institutional factors, ideological structures, and individual resources, and our theoretical framework involves a combination of these levels" (2018, 268). The analytical framework includes cultural and institutional dimensions, applied to a typical cultural sociology problem: the long-term change linked to the early gender *socialization* to peer and spousal interactions in later life, all within a wider context of *ideological norms*, regulatory systems and material constraints. The shift to the *cultural Coleman boat*—as they label it—has some key implications worth highlighting. First, in the *mainstream* conception of macro-micro mechanisms (Hedström 2005) the bottom left-hand corner of the Coleman boat implies *pre-formed agents*, without accounting for the cultural and symbolic mechanisms that *constitute* agents as such. The suggested happy wedding between analytical and cultural sociology invites us to consider the macro-micro mechanisms as prerequisites not just of actions, but also of *actors* and *intentions*. Like the micro-macro mechanisms are often—or in the more interesting cases—emergent, so the macro-micro of this kind are "immerging" (Barbera and Negri 2014). Immergence includes those properties that are *non-intentional* consequences of the interaction context in which agents are embedded. If the light

Filippo Barbera - 9781789906851

is on, I see the colours around me notwithstanding my intentions and those of the individual who turns on the light. This is in line with the Latin roots of the term *immergo*, which means disappearing by entering a medium, like the moon in the shadow of the sun. Like in the case of the micro-macro emergence, this doesn't exclude the generative role of actions (someone, after all, pressed the switch), but suggests that the *explanatory* added value lies in the properties of the context and in its key dimensions (intensity of the light, spaciousness of the place, spatial distance from the source of light and so on), not in the intentions of agents. Here arises the well-known dilemma between causal explanation and constitutive-compositional mechanisms that marks the debate in analytical sociology (Hedström and Bearman 2009). The light does not cause my possibility to see, it *constitutes* this possibility. If, on the contrary, immerging mechanisms are given a causal status, a logical contradiction occurs. If analytical sociology does not deem plausible the macro-macro *causal* mechanisms, why does it believe in the macro-micro ones? How can macro "generate" micro-effects? Shouldn't we be ready to accept that macro-micro and micro-macro relations are not characterized by generative causality, but just by constitutive-compositional effects that might be theoretically more important than the obsession for micro-foundation and intentionality? Doesn't the sociological explanation need both, after all? On this ground, the meaning of social action can be included in the macro-micro-macro framework even with regard to its most complex implications such as social identity and non-intentional processes. This would confirm the purely methodological nature of analytical sociological, which is compatible with a wide spectrum of substantive theories (for a statement, see Manzo 2010 and 2012). To include meaning-making mechanisms, analytical sociologists should be ready to bear an intellectual risk and distance themselves from the attempt to define a canon (Lizardo 2012), whose borders are designed chiefly to establish the approach as a powerful player in sociology's strategic action field.

REFERENCES

Abbott, A. (2004), *Methods of Discovery: Heuristics for the Social Sciences*. New York: Norton.

Abell, P. (1987), *The Syntax of Social Life*, Oxford: Clarendon Press.

Abell, P., T. Felin and N.J. Foss (2014), "Microfoundations of Social Theory: A Response to Jepperson and Meyer", *Sociologica*, **2**, https://doi.org/10.2383/78819.

Abrams, P. (1983), *Historical Sociology*. Ithaca, N.Y.: Cornell University Press.

Aksoy, O. and D. Gambetta (2016), "Behind the Veil: The Strategic Use of Religious Garb", *European Sociological Review*, **32** (6), pp. 792–806.

Barbera, F. (2011), "J.L. Borges, Reduction of Social Complexity", in C. Edling and J. Rydgren (eds), *Sociological Insights from Great Thinkers*, Santa Barbara: Prager, pp. 207–15.

Barbera, F. and N. Negri (2014), "Comment on Peter Abell, Teppo Felin and Nicolai J. Foss: Crypto-Rational Choice or Complex Mechanisms?" *Sociologica*, **2**, https://doi.org/10.2383/78820.

Barbera, F. and N. Negri (2015), "Rituals as Mechanisms", in G.L. Manzo (ed.), *Paradoxes, Mechanisms, Consequences: Essay in Honor of Mohamed Cherkaoui*, Oxford: Bardwell Press, pp. 115–39.

Bearman, P., J. Moody and K. Stovel (2004), "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks", *American Journal of Sociology*, **110** (1), pp. 44–91.

Becker, G. (1976), *The Economic Approach to Human Behaviour*, Chicago, IL: Chicago University Press.

Berger, P.L. and T. Luckmann (1967), *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*, New York: Anchor.

Bianchi, F., N. Casnici and F. Squazzoni (2018), "Solidarity as a Byproduct of Professional Collaboration: Social Support and Trust in a Coworking Space", *Social Networks*, **54**, July, pp. 61–72.

Blau, P. (1963), *Exchange and Power in Social Life*, New York: Wiley.

Boltanski, L. and L. Thévenot (2006), *On Justification. The Economies of Worth*, Princeton, NJ: Princeton University Press.

Filippo Barbera - 9781789906851

Boltanski, L. and T. Vitale (2006), "Una Sociologia Politica e Morale delle Contraddizioni", *Rassegna Italiana di Sociologia*, **30** (1), pp. 91–116.
Borges, J.L. (1998), *The Writing of God*, New York: Penguin.
Boudon, R. (1989), Subjective Rationality and the Explanation of Social Behaviour, *Rationality and Society*, **1** (2), pp. 173–96.
Boudon, R. (1993), "Toward a Synthetic Theory of Rationality", *International Studies in the Philosophy of Science*, **7** (1), pp. 5–19.
Boudon, R. (1998), "Social Mechanisms Without Black Boxes", in P. Hedström and R. Swedberg (eds), *Social Mechanisms*. Cambridge: Cambridge University Press, pp. 172–203.
Bourdieu, P. (1984), *Distinction: A Social Critique of the Judgement of Taste*, Cambridge, MA: Harvard University Press.
Carstensen, M.B. and V.A. Schmidt (2016), "Power Through, Over and in Ideas: Conceptualizing Ideational Power in Discursive Institutionalism", *Journal of European Public Policy*, **23** (3), pp. 318–37.
Centola, D. (2018), *How Behavior Spreads: The Science of Complex Contagions*, Princeton, NJ: Princeton University Press.
Centola, D. and M. Macy (2007), "Complex Contagions and the Weakness of Long Ties", *American Journal of Sociology* **113** (3), pp. 702–34.
de Vaan, M., D. Stark and B. Vedres (2015), "Game Changer: The Topology of Creativity", *American Journal of Sociology*, **120** (4), pp. 1144–94.
DiMaggio, P. (1997), "Culture and Cognition", *Annual Review of Sociology*, **23** (1), pp. 263–87.
DiMaggio, P. (2007), "Comment on John Goldthorpe/2", *Sociologic*a, **2**, https://doi.org/10.2383/24757.
Epstein, J. (2006), *Generative Social Science. Studies in Agent-Based Computational Modeling*, Princeton, NJ: Princeton University Press.
Elster, J. (1979), *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge: Cambridge University Press.
Esser, H. (1993), "The Rationality of Everyday Behaviour: A Rational Choice Reconstruction of the Theory of Action by Alfred Schutz", *Rationality and Society*, **1** (5), pp. 7–31.
Gambetta, D. (2009), "Signalling", in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press, pp. 168–94.
Goldberg, A. and S.K. Stein (2018), "Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation", *American Sociological Review*, **83** (5), pp. 897–932.
Goldthorpe, J.H. (2000), *On Sociology*, Oxford: Oxford University Press.
Hedström, P. and R. Swedberg (eds) (1998), *Social Mechanisms*, Cambridge: Cambridge University Press.
Hedström, P. (2005), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge and New York: Cambridge University Press.
Hedström, P. and L. Udéhn (2009), "Analytical sociology and theories of the middle range", *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press, pp. 25–49.
Hedström, P. and P. Bearman (eds) (2009), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press.
Jepperson, R. and J.W. Meyer (2011), "Multiple Levels of Analysis and the Limitations of Methodological Individualisms", *Sociological Theory*, **29**, pp. 54–73.
Kaufman, J. (2004), "Endogenous Explanation in the Sociology of Culture", *Annual Review of Sociology*, **30** (1), pp. 335–57.
Keuschnigg, M. and T. Wolbring (2015), "Disorder, Social Capital, and Norm Violation: Three Field Experiments on the Broken Windows Thesis", *Rationality and Society*, **27** (1), pp. 96–126.
Lawler, E.J. (2001), "An Affect Theory of Social Exchange", *American Journal of Sociology*, **107** (2), pp. 321–52.
Lawler, E.J. and J. Yoon (1998), "Network Structure and Emotion in Exchange Relations", *American Sociological Review*, **63** (6), pp. 871–94.
Lawler, E.J., S.R. Thye and J. Yoon (2000), "Emotion and Group Cohesion in Productive Exchange", *American Journal of Sociology*, **106** (3), pp. 616–57.
Lawler, E.J., S.R. Thye and J. Yoon (2008), "Social Exchange and Micro Social Order", *American Sociological Review*, **73** (4), pp. 519–42.
Lindenberg, S. (1993), "Alternatives, Frames, and Relative Prices: A Broader View of Rational Choice", *Acta Sociologica*, **36** (3), pp. 191–205.
Lindenberg, S. (2000), *The Extension of Rationality: Framing versus Cognitive Rationality*, reprinted from *L'Acteur et ses Raisons. Mélanges en l'honneur de Raymond Boudon*, a cura di J. Baechler et al., Paris: PUF, pp. 168–204.
Lizardo, O. (2006), "How Cultural Tastes Shape Personal Networks", *American Sociological Review*, **71** (5), pp. 778–807.
Lizardo, O. (2012), "Analytical Sociology's Superfluous Revolution", *Sociologica*, https://doi.org/10.2383/36902.
Manzo, G. (2010), "Analytical Sociology and its Critics", *European Journal of Sociology*, **51** (1), pp. 129–70.

Filippo Barbera - 9781789906851

Manzo, G. (2012), "Reason-based Explanations and Analytical Sociology. A rejoinder to Boudon", *Revue Européenne des Sciences Sociales*, **50** (2), pp. 36–65.
Manzo, G. (2013), "Is Rational Choice Theory still a Rational Choice of Theory?" *Social Science Information*, **52** (3), pp. 361–82.
Manzo, G. (ed.) (2014), *Analytical Sociology: Actions and Networks*, Chichester, UK: Wiley.
Milgram, S., L. Mann and S. Harter (1965), "The Lost Letter Technique: A Tool of Social Research", *Public Opinion Quarterly*, **29**, pp. 437–38.
Ostrom, E. (2005), *Understanding Institutional Diversity*, Princeton and Oxford: Princeton University Press.
Paluck, E.L. and H. Shepherd (2012), "The Salience of Social Referents: A Field Experiment on Collective Norms and Harassment Behavior in a School Social Network", *Journal of Personality and Social Psychology*, **103** (6), pp. 899–915, https://doi.org/10.1037/a0030015.
Parsons, T. (1937), *The Structure of Social Action*, Glencoe, IL: Free Press.
Pizzorno, A. (2006), "Rational Choice", in S. Turner and M. Risjord (eds), *Philosophy of Anthropology and Sociology*: *A Volume in the Handbook of the Philosophy of Science*, Volume 15, Dordrecht: Elsevier, pp. 373–95.
Rydgren, J. (2009), "Beliefs", in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press, pp. 73–93.
Sen, A.K. (1980), "Description as Choice", *Oxford Economic Papers*, **32** (3), pp. 353–69.
Sewell, W.H. (1996), "Historical Events as Transformations of Structures: Inventing Revolution at the Bastille", *Theory and Society*, **6**, pp. 841–81.
Sewell, W.H. (2005), *Logics of History*: *Social Theory and Social Transformation*, Chicago, IL: University of Chicago Press.
Simmons, C.H. and C. Zumpf (1983), "The Lost Letter Technique Revisited", *Journal of Applied Social Psychology*, **13**, pp. 510–14.
Stark, D. (2009), *The Sense of Dissonance: Accounts of Worth in Economic Life*, New York and London: Princeton University Press.
Stark, D. (2017), "Diversity and Worth: A Retrospective Account", *Sociologica*, **1**.
Sullivan, O., J. Gershuny and J.P. Robinson (1998), "Stalled or Uneven Gender Revolution? A Long-Term Processual Framework for Understanding Why Change is Slow", *Journal of Family Theory and Review*, **10**, pp. 263–79.
Swidler, A. (1986), "Culture in Action: Symbols and Strategies", *American Sociological Review*, **51** (2), pp. 273–86.
Udéhn, L. (2001), *Methodological Individualism*, London: Routledge.
Van de Berg, A. (1998), "Is Social Theory Too Grand for Social Mechanisms?", in P. Hedström and R. Swedberg (eds), *Social Mechanisms*, Cambridge: Cambridge University Press, pp. 204–37.
Vedres, B. and D. Stark (2010), "Structural Folds: Generative Disruption in Overlapping Groups", *American Journal of Sociology*, **115** (4), pp. 1150–90.
Wrong, D. (1961), "The Oversocialized Conception of Man in Modern Sociology", *American Sociological Review*, **26** (2), pp. 183–96.

Filippo Barbera - 9781789906851

8. Analytical sociology and pragmatism

Matthew Norton

Pragmatist sociology and analytical sociology have routinely been pitted and measured against each other (Abbott 2007; Gross 2009, 2013; Little 2012, 2014; Manzo 2010, 2014). This comparison and conflictual framing are tempting. Both pragmatist and analytical sociology share an interest in mechanisms-based approaches to sociological explanation and have a similar enough conception of mechanisms to allow for easy comparison and contrast. Furthermore, this exercise has been valuable because for each school of thought, having an approach with a similar starting point but a different emphasis and set of assumptions has been useful for clarifying what each is up to and why. On the other hand, the points of common interest between analytical sociology and a pragmatist mechanisms approach have usually been taken to highlight the stark differences just beneath the surface of these schools of thought on matters of causal logic, levels of analysis, individualism, and the centrality and theory of meaning, among others. The premise of this chapter, however, is that a series of clarifications and developments by analytical sociologists on matters related to methodological individualism, the Analytical Sociology (AS) theory of action, the claim that AS can best be understood not in reference to a specified causal logic but rather as "a 'syntax' for explanation" (Manzo 2010: 162) and through clarifications and developments of the mechanism concept have resulted in a rather more accessible conception of analytical sociology for those adopting pragmatist approaches to sociological analysis. These developments make possible, and valuable, a different way of thinking about the relationship between pragmatism and a pragmatist approach to mechanismal explanation and analytical sociology. This chapter will explore the possibility and value of doing analytical sociology in a pragmatist mode.

After first noting several of the commonalities that have brought AS and pragmatist approaches into frequent contact, the chapter will consider four points that help to define AS in the pragmatist mode. It first examines methodological individualism as a central focus of the pragmatist/AS divide and how more recent definitions of AS's so-called structural individualism can be developed in a manner that is consistent with a pragmatist focus on multiple contextualized situations. Second and third, it will adopt a similar approach to examining how theories of action and causal logics that have once been taken to mark the divisions between these approaches can, given important clarifications of the AS school of thought, be adapted to a pragmatist view. Finally it considers the analytical foci of AS in the pragmatist mode, describing how the conciliatory theorizing of the preceding sections opens the way for an analytical sociology focused on mechanisms that highlight typical pragmatist concerns about the social world and strengths in analyzing it that ultimately enrich the analytical sociological approach and broaden its range in important ways.

1. POINTS OF ALIGNMENT

The value of the comparison between analytical sociology and a theory of mechanisms based on pragmatist social theory depends in part on the similarities between these approaches. Most obviously, both approaches adopt the concept of mechanism as central to their

Matthew Norton - 9781789906851

approaches to sociological explanation (Gross 2009). There have of course been many definitions of mechanism advanced in analytical sociology, each with a slightly different emphasis or inflection. The risk of this approach is that any critique aimed at a specific definition can be unproductively stymied by shifting to another of the available definitions. For that reason I focus here on definitions that are mainstream, but also that are most amenable to understanding the commonalities between AS and pragmatist social theory. In one of the ur texts of contemporary analytical sociology Hedström (2005) writes that "a social mechanism, as defined here, is a constellation of entities and activities that are linked to one another in such a way that they regularly bring about a particular type of outcome" p. 11). More generally, Hedström and Bearman (2009) argue that a good account will describe how the mechanisms in play generate the outcome to be explained (p. 6). Hedström and Ylikoski (2010) have also influentially described mechanism-centered explanation as focusing on the "cogs and wheels of the causal process through which the outcome to be explained was brought about" that "generate aggregate outcomes" (p. 50)

The analysis of the genesis of outcomes of interest has meant that analytical sociology focuses on building up chains of causation and specifying the elements of those chains and their relations (Gross 2018; Abbott 2007). In doing so AS aims to specify and clarify the component objects, actions, relationships, and other factors that combine to constitute mechanisms that can be used to describe the generation of outcomes of interest. In doing so it aims to avoid placing "black boxes" athwart the explanation and thus breaking the enchained sequences of mechanisms and effects that generate outcomes and are the hallmark of the AS approach to sociological explanation. In practice this has meant a focus on identifying causal factors at lower levels of analysis than the outcome of interest and exploring how those mechanisms interact, combine, colligate, and aggregate to generate higher level outcomes, and vice versa: how structural factors are translated into the concrete contexts of social action.

This approach coincides neatly with one of the founding imperatives of pragmatist social theory: the pragmatic maxim. As formulated by Peirce (1978), the pragmatic maxim held that in order to explain something the analyst should "consider what effects, that might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object p. 402). This pragmatist emphasis on explanation through a focus on effects is the foundation for linking pragmatist approaches in sociology to contemporary analytical sociology. Both look to the production of effects in order to explain something as opposed to approaches that instead seek to explain, for instance, via constant conjunction or deduction. And both analytical sociologists and pragmatist social theorists seem to be in general agreement with an approach to explanation that seeks to account for the generation of effects of interest.

This general amenability of pragmatist social theory and analytical sociology can account for the further concordance of these approaches on a similar style of explanation that focuses on causal accounts anchored in specific, concrete, and clear social occurrences and patterns and the consequences that they generate. The pragmatic maxim cited comes from Peirce's essay "How to Make Our Ideas Clear," and this focus on clarity as a central stylistic and even epistemic goal of explanation is likewise clearly echoed in foundational works of analytical sociology (Hedström 2009; Hedström and Bearman 2009: 3–4; Manzo 2014: 10–11) (see also Little's chapter on analytical philosophy and sociology on this volume). Both abjure underspecified and overly abstract explanations that fail to account

Matthew Norton - 9781789906851

for causation through specified generative sequences and fail to anchor these causal models in the nitty-gritty actuality of social life where the causal force of all mechanisms is ultimately paid off.

Both analytical sociologists and pragmatists also seem to be in agreement on the importance of developing explanations of social phenomena up to the macro level that not only remain consistent with sociological knowledge and theory about the character of human social life at more constitutive levels but that actively seek to elucidate their explanation in terms of the sociological factors at these more constitutive levels that generate trans-situational, multi-faceted phenomena occurring over longer temporal scales and wider spatial scales than any discrete action or situation alone and which are constituted by the aggregation of action sequences.

It is this concordance that has helped to put pragmatist social theory and analytical sociology in conversation with one another, for each has been represented as attempting to provide a roadmap for good explanatory practice (Hedström and Bearman 2009; Little 2012, 2014; Timmermans and Tavory 2012), and each tradition has generated significant discussion of what the mechanism concept means for sociology as a matter of both method and theory. It is in the context of the dominance of the middle-range in much of contemporary sociology that these two approaches have become somewhat prominent, for each offers a playbook for how to engage in middle-range–focused sociological explanation in a manner that is consistent with a defensible causal epistemology itself rooted in a defensible, if not fully developed, social ontology.

As the several salvos of critique and restatement between pragmatist proponents of mechanism-based sociology and analytical sociologists suggests, however, these approaches are not in easy agreement. Those differences can be reconciled, however, in the context of broader and more principled articulations of the AS project.

2. THE SHORT HOP FROM METHODOLOGICAL INDIVIDUALISM TO METHODOLOGICAL LOCALISM

For a sociological method focused on providing clearly specified mechanismal explanations, the attractions of individualism are clear. Individualist approaches offer an easily graspable, intuitively bounded nexus for constructing mechanismal explanations by focusing on a unit of analysis, the individual, that plays an obvious central role in action and coordination. As rational choice approaches demonstrate (Kiser 1994; Kiser and Hechter 1998), it is possible to generate explanations for even quite complex social facts and events by building from an abstract, simplified conception of the individual. Analytical sociologists have resisted the rational choice flavor of individualism as a foundation for mechanisms-based sociological explanation in general, however. The (intentional) reductivism and toleration of implausible micro-level models for the sake of building explanations from clear and simply stated theoretical premises that is the hallmark of rational choice approaches imposes too high an analytical cost for an approach like AS that undertakes to build generative models that are empirically valid every step of the way and that aim to account for the actual processes that generate outcomes of social life.

Analytical sociologists have, however, emphasized the individual as the preeminent unit for creating sociological explanations that are epistemologically defensible because they are

Matthew Norton - 9781789906851

rooted in a level of action that is indisputably concrete. Hedström's (2005) observation, that "the causal efficacy of actions would be readily seen if we were able to press a pause button that suddenly froze all individuals and prevented them from performing any further actions. All social processes would then come to an immediate halt" (p. 28), speaks to the attraction of the individual for AS. AS implements its version of individualism through the concept of "structural individualism." Structural individualist doctrine holds that "social facts should be explained as the intended or unintended outcomes of individuals' actions. Structural individualism differs from traditional methodological individualism in attributing substantial explanatory importance to the social structures in which individuals are embedded" (Hedström and Bearman 2009: 5). It is worth noting that there is significant ambiguity in how to understand this and similar pronouncements. On the one hand it can be read as individualism with an epiphenomenal gloss of institution and culture talk, easily hand waved away when the time comes to get to the actual models. On the other hand it presents a view that can barely be called a real individualism at all as it brings non-individual levels of analysis that exhibit significant emergent properties into the explanatory model (Erikson 2017).

The AS emphasis on individualism has typically been at the front of critiques from a pragmatist direction (Gross 2009). The gist of these critiques is that because of its insistence on the importance of being able to reduce mechanisms-based explanations to the actions and even the motivations of individuals, structural individualism diminishes the analytical fidelity with which it can represent the importance of the emergent properties of situations and the powerful contextualization that meso- and macro-level social facts and forms impose on individuals through the ways that they structure situations. This argument holds that structural individualism is a relatively transparent attempt to have one's cake and eat it by insisting on the importance of context, interaction, and structure, but adopting a theory of the individual that does not match these protestations of concern.

In response to these criticisms (and I should note at this point that it is not my intent in this chapter to wade into whether these debates have clarified what was always the AS position or if they have pushed AS into developing more nuanced theoretical frameworks) analytical sociologists have clarified their commitment to both parts of the so-called structural individualism that is the core tenet of its micro-level theory of the social actor. Manzo (2014), for instance, insists that both "macro-and meso-level entities and properties are regarded as causally important by analytical sociologists. From an explanatory point of view, the only requirement is that it should be possible to indicate at least one micro-level element through which the macro/meso-to-micro causal effect is generated—no matter if consciously or unconsciously from the point of view of the micro-level entity at hand" (pp. 19–20). He goes on to emphasize that these macro- and micro-level factors should be understood as exhibiting emergent properties with respect to individual-level aspects of explanations. Manzo's clarification is important because its implication is that for AS macro and meso elements of explanations *cannot* be reduced to individuals and must instead be built into sophisticated, multi-level models that account for emergent properties as well as dynamics of aggregation and structuration. Of course it also holds the opposite—that AS explanations cannot be divorced from individuals either, but as Lizardo (2012) writes this is a stipulation it is hard to imagine many sociologists disagreeing with (p. 5).

For AS in the pragmatist mode, Manzo's clarification of structural individualism opens the door to a way of doing AS that is aligned with the typically pragmatist micro-foundational focus on situations and interaction. Indeed, Manzo's restatement of the structural individual-

Matthew Norton - 9781789906851

ism concept with an emphasis on the reality of the commitment to the structure part raises the question of whether AS needs to identify itself with methodological individualism any more strongly than any other way of doing sociology that notices that "people doing stuff" is at some level germane to the analysis.

Pragmatist approaches to mechanisms share with AS a desire for concrete foundations on which to build explanations and turn to situations rather than individuals to provide that clarity and specificity. As a general matter it is not clear that individuals should be taken to represent a more micro level of analysis than situations. To turn the tables on Hedström's claims regarding the intuitive plausibility of individualism, it is equally true that were we to magically pause all situations or interactions "all social processes," would likewise, "come to an immediate halt" (Abbott 2001; Hedström 2005: 28). While perhaps counterintuitive—don't individuals constitute situations?—doubts about the clear preferability of individuals vs. situations as microfoundational levels of analysis can be supported by observing that crucial aspects of social life do not occur, or do not occur in the ways that matter for most sociological purposes, at the level of individuals. To put it more bluntly, one of the purposes of analytical concepts is to carve social life at its joints in ways that are conducive to constructing valid generative models and sociological explanations. But from a pragmatist perspective, the individual isn't a joint. Rather, the individual only exists for analytical purposes as embedded in and constituted by a dynamic relationship with the situations that they find themselves in. As Rawls (1987) writes, "The social self needs to be continually achieved in and through interaction" (p. 136). Carving the individual from the situation, as intuitively attractive as that might seem, from a pragmatist perspective cuts across the very circuits of translation and influence, the very dynamics of embeddedness, that provide the microfoundational basis on which AS in the pragmatist mode will need to be built. It is an analytical maneuver that runs the risk of the very theoretical abstraction that AS in other contexts abjures because it posits as foundational a highly abstract and entirely implausible concept of the non-situated individual.

Little has proposed a microfoundational concept that is similar to structural individualism in some respects but that more unambiguously corresponds to a pragmatist view of this microfoundational question and that may be more suitable than structural individualism as an orienting analytical theory for AS in the pragmatist mode. Methodological localism, Little (2007) writes, takes as its centerpiece the "socially constructed and socially situated individual, who lives, acts, and develops within a set of local social relationships, institutions, norms, and rules" (see also Little 2008). The advantage of this view is that it makes the relationship between individuals, situations, and other situation elements, including structural factors that are real and present circumstances of the local context of action, the key microfoundational unit for sociological analysis (Schneiderhan 2011: 595).

Taking situations, and more specifically the relations and patterns of dynamic circulation that constitute them as the central microfoundational focus in place of individual entities with their properties offers a microfoundational account that is consistent with the goals of AS expressed by Manzo but that is far more amenable to constructing explanations along pragmatist lines and more clearly foreground characteristically pragmatist analytical foci. Interaction with other people and with the physical environment, for instance, are prominent examples of social action that cannot be reduced to the individual and nonetheless are microfoundational elements of analysis. Pragmatist microfoundations insist on accounting for the primordial sociality and corporeality of individuals (Joas and Beckert 2002), which is to say that individuals and individual action cannot be usefully conceptualized outside of the situational context

Matthew Norton - 9781789906851

within which these constitutive interactions happen. The translation of signs from individuals to social environments and back likewise cannot be mapped from an approach that cuts the micro-level off by defining it as pertaining to the individual. Furthermore, sociology has an excellent set of methodological approaches, ranging from cultural sociological analysis to symbolic interactionism and comparative historical process tracing, for assembling situations into patterned chains of effects. A theory of the structure of situations (Norton 2014) is likewise critical for pragmatist accounts of social action as well as accounts of the dynamics of aggregation and influence that define relationships between the macro and micro levels of analysis and that are easier to account for in methodological localist terms that decline to separate the individual from the situation. In this view, situations are the key mechanism (or meta-mechanism) used to account for the translation of macro structures into experientially real constraints and affordances that shape the possibilities and probabilities for different lines of action. The point is not that these cannot be considered on structural individualist grounds, but that doing so is less elegant than doing so on the basis of methodological localist grounds and in any case is not necessary from the perspective of AS.

For the purposes of the argument of this chapter, though, the main point is that the social ontology of AS is broadly enough conceived to incorporate both structural individualism and methodological localism and furthermore both approaches are well-conceptualized enough that they can serve as platforms for structuring and organizing different lines of AS work. There is space in AS for a social ontology that enables AS in the pragmatist mode.

3. ACTION

Given their commitment to constructing explanations anchored in concrete and specific circumstances of social life, AS and pragmatist theories of social mechanisms are aligned in the central importance of a theory of action. It is the specific content of action theory that has provided grist to accounts focusing on the conflicts between these approaches (Gross 2009; Manzo 2010). For AS the most prominent action theory has been Hedström's DBO approach. DBO, like structural individualism, parts ways with rational choice theory despite some shared premises and concerns. DBO, though, argues that to understand individual action we should look to the sociologically embedded genesis of actors' desires and beliefs rather than imputing them on the basis of a general theory of rationality. This largely anodyne theory (Lizardo 2012) has nonetheless been a flashpoint for critics, refuting the idea that DBO, at least in its initial statement, would provide an uncontroversial theory of action. Reasons that DBO has turned out not to be widely acceptable as a generic sociological theory of action are useful to relate here only insofar as they help to establish the parameters for a theory of action for AS in the pragmatist mode.

While DBO theory suggests desires as a more inclusive alternative to the doctrinaire rational choice emphasis on instrumental utility maximization, critics have warned that for sociological purposes we cannot take desires to be established prior to and apart from the circumstances of the analysis (Gross 2009: 365). They must be treated as an endogenous dimension of the problem of analysis itself, not as established prior to the action situation to be explained. Beliefs are DBO theory's implementation of culture on the level of action (Rydgren 2009). As with desires, however, critics have warned against an overly individualistic and static concept of beliefs as something that an individual brings to the action situation, rather than a part of the

Matthew Norton - 9781789906851

analysis of the situation and importantly something that emerges from the interaction between people and situations. Further, the notion of beliefs needs to be clearly defined and specified so as not to become the sort of vague implication that analytical sociology claims to guard against. This is especially important to note for approaches like AS in the pragmatist mode in which culture and meaning, "the semiotic dimension of human social practice" (Sewell 2005: 164), are at the heart of the analytical project. In place of focusing exclusively on culture concepts best attuned to individual minds and selves, pragmatist approaches need to make use of the much more robust and diverse conceptual repertoire developed under the auspices of pragmatist, symbolic interactionist, and cultural sociology for the analysis of culture and meaning at the cognitive, individual, situational, meso, and macro levels of analysis (e.g. Alexander 2004; Eliasoph and Lichterman 2003; Fine 1991; Norton 2019; Reed 2013; Tavory and Timmermans 2013; Wagner-Pacifici 2010).

Recent defenses against these critiques, however, have insisted on the modularity of DBO as AS's preferred theory of action. In his response to pragmatist and other critics of DBO Manzo declines to argue for the merits of the more restricted view of DBO theory ascribed to it by its critics, instead arguing that on the multiple points where the differences between DBO and pragmatist theories of action have been identified there is no conflict at all. He writes of DBO theory:

> 1) it makes no assumptions on the specific form of actors' rationality; 2) actors' ends and means may arise from their action rather than precede it; 3) actors' beliefs and desires are conceived as explananda, not as exogenous elements of the problem; 4) actors' beliefs and desires are not conceived as being invariably clear, transparent and consistent; 5) the cognitive and relational mechanisms responsible for actors' beliefs and desires are conceptualized as being able to operate unconsciously; 6) actors' beliefs and desires are inseparable from actors' perceptions of the symbolic context in which they act.(Manzo 2010: 157)

Accepting this argument at face value, which I do, establishes that DBO theory is just one possible theory of action that has some affinity for the AS approach to explanation but is hardly the only theory of action that might be useful. This leaves the particulars of the AS theory of action wide open to pragmatist approaches.

Indeed, pragmatist approaches are already closely aligned with the metamethodological principles that provided encouragement for the uptake of DBO theory as the main AS theory of action in that they are concrete, clear, and specific. Nonetheless, pragmatist action theory differs significantly from DBO theory and provides AS in the pragmatist mode a very different valence from approaches focused on DBO or similar action theories. For AS purposes, a pragmatist theory of action should be understood as consistent with the foundation of "methodological localism" described in the preceding section. One of the core concepts of pragmatist action theories is that beginning with the individual is a mistake because action is fundamentally structured by the relationship between people and situations. Attempting to clarify action-theoretic mechanisms by the abstract extraction of the individual from the situations where action actually happens and then theorizing on this basis, which the DBO approach invites if not insists on, is to cut *the* crucial mechanismal connections for effectively theorizing action according to a pragmatist view. As Joas writes, "goals are not externally set but emerge in the action process itself in a reciprocal interaction between means and goals … as part of the action process that is inherently connected to the situational context" (Joas and Beckert 2002: 273).

Matthew Norton - 9781789906851

What this means theoretically is that motives emerge through the course of experiencing structured situations and interactions. Similarly, opportunities should also be understood as endogenous to and embedded in situations, observations of unfolding situational circumstances that provide actors with possible lines of action leading to "ends in view" (Dewey 1958; Joas and Beckert 2002: 273). Action is thus conceptualized in a way that remains concrete, specific, and clear, but is much less amenable to the individual frames of reference that DBO theory emphasizes. Likewise this approach shifts the focus from entities with properties to both entities and their properties as relationally constructed, constituted, and reciprocally ratified through interaction in the course of a situation (or multiple situations that share features). Entities and properties are not conceptualized prior to the action context but are rather made present realities of the structure of situations through action and then in turn structure further and future action in that context. From the vantage point of pragmatist action theory, then, motivations are better understood as emerging out of the habitual or creative responses of actors to situations, and the analysis of situations can be understood as an effort to understand how and why situations call forth certain actions and lines of action from the people experiencing them (Camic 1986; Gross 2009; Joas 1996; Silver 2011).

Thinking about the courses of action that particular situations, whether routine or problem, constrain or afford also provides an important way to think about meaning as a part of situation structure and individual experience in a manner that goes beyond the underspecified and static notion of culture implied by the notion of beliefs. According to the semiotic conception of culture preferred by pragmatist theory, culture refers to the circulation and translation of signs into and between situations in a manner that shapes those situations as environments of action (Norton 2019). Peirce (1978) emphasizes this tight linkage between meaning and action when he writes that "the whole function of thought is to produce habits of action" (p. 400). The meaningful component of action in this view is not simply what someone thinks (or believes, or how they cognitively process situational information), though this is certainly part of the picture, and more about what representations are present (or absent), how they bring situational structures into existence, how they are translated into individual experience, and how they are connected to past and future situations in ways that create meso- and macro-cultural structures that can be understood as key mechanisms in shaping the structure of other situations. They are public culture in the sense that they are manifest as situational realities more so and more importantly than they are manifest in the interstices of individual minds (Lizardo 2017).

This is a different basis for action theory than DBO obviously, but given Manzo's clarification it is nonetheless consistent with the purpose of DBO for AS: to provide a concrete, clear, and explicit theory of action that can serve as a fundamental, indeed perhaps the fundamental, element in a mechanismal approach to sociological explanation. Desires, in the pragmatist mode, are reconceptualized as motivations in a pragmatist theory of action and conceptualized as emergent from structured, situational interaction. Beliefs recede from the theoretical framework in this view and are replaced by a more involved, but still well-theorized, specific, and clear set of concepts to describe and model the orientation of action to social meaning manifest in ways dependent on situational specifics and in ways that are shaped by large-scale structures and patterns of meaning. Opportunities refer to the structure of situations, and especially the ways that they are structured by both contingent situation-specific meanings and by situational manifestations of meaning structures that occur over longer temporal

Matthew Norton - 9781789906851

and wider spatial scales than a single situation. Finally, there is an overall greater degree of attention paid by this pragmatist theory of action to the play of habitual and creative action as triggered by the specific constellations of motive, meaning, and structure that comprise a given situation as well as to the ways in which meaning operates to link situations to broader structures of meaning and power as well as to the ways that situations present possible and likely future consequences that are likewise part of actors' experiences of the meaning of the situation.

4. CAUSALITY AND EXPLANATION

A similar framework to that deployed in the previous sections pertains to questions of AS's causal theory and its compatibility with those commonly used in pragmatist sociology. AS has at times looked to a theory of causality based on "counterfactual dependency" (Hedström and Ylikoski 2010: 54) and thus, to at least some degree, a theory of causality focused on the independence of the mechanisms that contribute to an explanation (Knight and Reed 2019). Counterfactual dependency, however, is not a programmatic requirement of AS so much as one way to think about the relationships between the mechanisms that contribute to a particular outcome that will be useful in some analytical contexts more than others. Indeed, Hedström and Ylikoski (2010) specifically tout as an advantage that "the mechanism-based account of explanation is not wedded to a specific theory of causation" (p. 53). Rather, the core AS commitment is to generative models as explanations, and such models can be constructed in valid ways on the basis of different conceptions of causality. So while an account focused on counterfactual dependency and independent, "modular" mechanisms (Knight and Reed 2019) may be valid for some explanations, for others mechanisms underwritten by different concepts of causality will yield better generative models.

The concept of causality that runs through the work of contemporary sociological pragmatism is, generally speaking, probabilistic, concerned with the interdependency of mechanisms and the sensitivity of causation to the particulars of situations, and skeptical of the intersituational stability of the entities and properties, such as belief bearing individuals, that other approaches to mechanismal explanation often take for granted (Norton 2019). For pragmatist explanations it is often necessary to think about mechanisms as generating outcomes in conjunctural ways, and often in complex conjunctural ways (Steinmetz 1998: 178). This means that to explain a given outcome it will often be necessary to model multiple interacting mechanisms that operate differently in the specific conjunctures that occur in the empirical context examined than they would in other conjunctures. This leads to a potentially uncomfortable but necessary concept of non-deterministic mechanisms that can only be properly modeled in consort and in specific configurations. Pragmatist explanations, as focused as they are on a semiotic theory of meaning, are also rarely best pursued on the basis of independent mechanism. This is because semiosis involves the pervasive presence of relationally defined, intrinsically interdependent signs and semiotic structures. As Peirce (1978) writes, "This universe is perfused with signs" (p. 448), a perspective that powerfully shapes the entirety of the pragmatist approach and its focus on meaning. The representations that are generated in the situations that actors experience, therefore, are necessarily contingent in their effects on interdependent networks of signs that are incorporated into the situational circulation and manifestation of meaning through the interpretive experiences of situational

Matthew Norton - 9781789906851

actors. As Tavory and Timmermans argue, a particular configuration of meaning introduced into dissimilar situations can lead to quite different outcomes. That configuration of meaning is correctly understood as contributing in an identifiable way to the genesis of both outcomes. Its effects, however, depend on the conjunctures of other mechanisms that they are part of and that specify the meanings that actors actually experience and respond to (Tavory and Timmermans 2013).

It is important to note that this probabilistic, conjunctural, semiotic, interdependent approach to explanation is nonetheless consistent with the goal of creating explanations that identify specific chains of mechanisms that generate outcomes of interest. It is just that the generation of outcomes, from a pragmatist view, is often a conjunctural, probabilistic, and interdependent matter that cannot be adequately or accurately accounted for in terms of independent, fully modular mechanisms (Knight and Reed 2019). That ontological complexity does not call for throwing up one's hands at the prospect of generating specific, clear mechanismal explanations, but does suggest that certain ways of doing so that are conjunctural, probabilistic, and interdependent are preferable, at least for the types of outcomes that pragmatist analytical sociology tends to focus on. Fortunately, this way of thinking about causality and explanation appears to be fully consistent with an approach to AS that understands mechanisms as operating through the highly context-dependent dynamics of situational conjunctures of meaning.

5. METHODS

One of the ongoing lines of argument in this chapter has been that analytical sociology is a far more inclusive platform for sociological analysis than is often recognized (though see Lizardo 2012), particularly when more recent clarifications and restatements of the core elements of AS are taken into account. If this argument is correct, then different ways of doing AS should be expected to emerge with different methods, different questions, and different analytical foci.

Approaches that may make sense in the context of other modalities of AS will be superseded in pragmatist approaches because of their different foci and causal theories. For instance, Manzo (2014) has argued for the centrality (though not the exclusivity) of agent-based models as a strong approach for developing generative explanatory models. In the pragmatist mode of AS, however, agent-based models have little to offer. Pragmatist analytical sociology is more commonly going to be concerned with the close observation of specific interaction situations along with the theoretical and empirical reconstruction of the various mechanisms that structure those situations, including their patterns of motivation and action, as well as the mechanisms of aggregation and influence that enchain situations and macro structures together for purposes of explanation (Fine 1991; Gross 2018). Pragmatist analytical sociology, following pragmatist sociology more generally, must therefore focus on methods that foreground the observation and analysis of situations of important types as well as *sui generis* situations that play consequential roles in the transformation of structures (Sewell 1996).

Historical (Vaughan 2004) and in-person ethnography (Tavory and Timmermans 2013) are good examples of pragmatist AS methods. These approaches focus on analyzing specific interactions in detail and provide the empirical foundation necessary to begin building an understanding of the semiotic structure of the types of situations that help to generate the outcome of interest. Those structural investigations themselves are a further foundational

Matthew Norton - 9781789906851

element of AS in the pragmatist mode as the researcher looks to reconstruct the circuits of meaning that recur and concatenate in ways that generate the outcomes of interest. For the same reason, case studies (Gibson 2011; Norton 2014) will often serve as a valuable tool for this way of doing analytical sociology because they allow for the close observation and careful reconstruction of the circulation of meaning and its causal consequences in a delimited circumstance that may then be able to be extended into a middle-range mechanismal theory. By examining these dynamics in the context of the specificities of a delimited case the researcher is better able to understand the interaction between meaning and action as well as the ways that these situation-level dynamics combine to produce larger patterns of meaning as well as more macro mechanisms such as institutions. Both of these approaches, it should be noted, are closely connected to and can draw on a range of techniques from more general methodological traditions, including phenomenology and symbolic interactionism more generally.

Process tracing is another important tool in the effort to examine specific mechanismal links between situations and situation-level dynamics of meaning and more macro social structures and outcomes (Mahoney 2012). In a general sense, process tracing involves identifying occurrences that are connected temporally through enchained mechanisms by which earlier events lead to or powerfully shape later events. While often used in macro-historical analysis, this same method can be useful for identifying and modeling any processes that occur over time. Process tracing methods serve as one of the most powerful ways for identifying, analyzing, and assembling generative models from the sorts of mechanisms that pragmatist analytical sociology specializes in. They are also a valuable way to address concerns that approaches to explanations from a probabilistic, conjectural, and highly contextualized perspective are liable to the production of "just so stories" (Bearman 2012: 2). The systematic analysis of causal sequences in close empirical detail offers a valuable corrective to underspecified processual accounts. Between-case comparisons offer another methodology of great complementary importance to process tracing in assembling and validating typologies of mechanisms that operate in definable ways even in complex situational contexts (Tavory and Timmermans 2013).

Pragmatist AS will also, of course, focus on methods that allow for the scientific and systematic investigation of the semiotic dynamics that structure social life and through which much of social life transpires. Meaning is at the core of the pragmatist modality of AS. Thus, a wide range of tools drawn from cultural sociology, including hermeneutic approaches, structural analysis, and the analysis of performativity, just to name a few, may all find a place in specific pragmatist AS lines of inquiry. Semiotic analysis in the Peircean tradition with its emphasis on the situated consequences of meaning-making likewise has much to offer. Importantly, semiotic analysis in this view does not inhere in the analysis of abstractions. Indeed, to the contrary. Approaches like those focused on the material manifestations of culture as a corporeal reality that confronts people in the situations that they experience provide powerful frameworks of analysis (McDonnell 2010). Semiotic approaches are especially important as they attempt to put meat on the bones of the claim that culture is a matter of causal significance and not just a matter of context with unspecified contributions to explanation understood in a generative way. Indeed, these semiotically oriented methods from cultural and pragmatist sociology that aim for a systematic account for the causality of culture and meaning in a way that is robust to its many varied manifestations is one of the most important lessons from pragmatist AS that needs a wider airing amongst adherents of AS more generally.

Matthew Norton - 9781789906851

6. PRAGMATIST FOCI

AS in the pragmatist mode differs from other modalities of AS, focusing on typically pragmatist concerns of a semiotically inclined analysis of meaning as an aspect of situation structure, interaction dynamics, and in general one of the most important mechanisms, or more precisely genres of mechanisms, for sociological explanation. This is not to suggest that the rest of AS ignores meaning, but pragmatist approaches, aligned in this regard with the concerns, theories, and methods of cultural sociology, engage in a more robust theorization and subsequent analysis of the semiotic dimensions of social life than an approach that takes the relatively simple cultural theory behind DBO or a similar individual-centered theory as their foundation for thinking about meaning. This focus on semiotic dynamics is rooted in pragmatist theory. Specifically, Peirce's theory of the sign provides a useful sketch of the main elements of this analytical tradition, focusing on the signature series of translations that comprise semiosis, from mental or physical objects to signifiers to the effects of signification all incorporated into a "continuous temporal process of interpretation" (Rochberg-Halton 1986: 46).

Semiotic analysis of this sort, and more generally the pragmatist and cultural sociological focus on meaning as one of the fundamental sociological mechanisms, has a role to play in analyzing specific situations and their dynamics. It is also central to understanding how situations and situational outcomes can become aggregated and colligated in a manner that amounts to the emergence of meaning structures and other elements of social structure. Those aggregation dynamics taking us from patterned interactions to institutions and structures that transcend individual situations and take on causal significance of their own as well as through their potential structuring of subsequent situations. Likewise this modality of AS focuses on the ways that actors translate the realities of social structures such as institutions or distributional facts into situations in a way that makes them a real structural circumstance for that situation. This focus, again, is especially concerned with meaning as a key category of mechanisms for understanding this translation from the macro to the micro level of the situation. This, to be sure, is itself a methodologically complex operation and likely requires an all-hands-on-deck approach when it comes to identifying appropriate methods. Agent-based models are one technique for doing so and that AS practitioners already draw on and that could be used to study these aggregation processes (Manzo 2020). Other methods that offer ways to understand these aggregating processes and mechanisms include structuralist and structural hermeneutic approaches (Alexander and Smith 1993; Norton 2011, 2014) and process tracing methods drawn from historical sociology that attempt to understand the specific chains of causation that connect concrete interactions with outcomes of broader significance (Gross 2018; Mahoney 2012; Sewell 1996).

A further concern that AS in the pragmatist mode should exhibit, corresponding closely with these micro-to-macro and macro-to-micro aggregation and translation dynamics, is with the role of forming causes in sociological explanation. Hirschman and Reed (2014) have suggested a distinction between "forcing causes" and "forming causes." The former loosely corresponds with the sorts of mechanisms that are AS's traditional focus. For example, discrimination by hiring managers is an important forcing cause of high levels of unemployment and the subsequent integration problems experienced by ex-felons in the United States following release from jail or prison. While this could be part of a pragmatist AS analysis of questions related to recidivism, for example, the felon construct would also play a central role in that analysis. Furthermore, it would not play a role as context, part of opportunity structure,

Matthew Norton - 9781789906851

or a background concern, but as an important cause unto itself. The difference between these causes, to the extent that one is necessary, is that an analysis of forming causes recognizes the semiotic webs produced and reproduced through interactions and translated into the concrete contexts of many of the situations that contribute to the phenomenon of interest should themselves be recognized as playing a fundamentally important causal role by bringing into being within and across situations the realities of meaning that determine the flow of action. Discrimination against felons obviously could not exist without felons, but more importantly the semiotic dynamics that continually form and reform the realities that determine the meanings—in Mead's (1934) sense, "What you are going to do about it?" (p. 49)—of "felon" as a social category.

CONCLUSION

The mechanism metaphor as it is used in analytical sociology trades on the promise of clarity. It famously enjoins us to find the cogs and wheels of the causal process and to report back on them with fidelity and specificity to achieve a more precise causal science. The second part of the analytical sociology premise is that such clarity can be achieved without resorting to implausibly simplistic, reductionist, or mechanistic models of social action. The extent to which it succeeds in achieving both of these goals at once is one principle measure of the analytical sociology program. It is not a straightforward measure, however, couched as it is in the ambiguous terms "clarity" and "implausibly." To get a sense of its progress in delivering on these ambiguous goals, and to therefore gain a sense of where the program is doing well, where it has run into trouble, and where it can and should go from here it is helpful for the practitioners of analytical sociology to look to close comparators. The pragmatist theory of mechanisms offers just such a valuable comparator. While the traditions differ on significant points and it may be temptingly pluralistic to let each get on with it, this would be a lost opportunity. The analytical sociology and the pragmatist theory of mechanisms share some important premises (generating explanations of social phenomena up to the macro level that remain consistent with sociological knowledge and theory about the character of human social life at more constitutive levels) and core metaphors (mechanisms). They are hardly in easy accord though. The pragmatist theory of mechanisms insists that the relational, processual, contextual, semiotic, and intrinsically interdependent and interactive character of social life sets an epistemic bound on how finely we can carve social life into its component parts before we find that what we have left is inadequate to the task of adequate sociological explanation, and that this bound is rather less micro, and that an adequate theory of social action is more relational and interdependent, than what is generally preferred by AS. The premise of this chapter, however, has been that there is daylight between the premises and usual practices of AS and that an AS in the pragmatist mode is not only feasible but has much to recommend it. Contrasting the approaches continues to have value, but so may integrating them.

REFERENCES

Abbott, Andrew (2001), *Time Matters: On Theory and Method*. Chicago: University of Chicago Press.
Abbott, Andrew (2007), "Mechanisms and Relations", *Sociologica* **2**: 1–22.
Alexander, Jeffrey C. (2004), "Cultural Pragmatics: Social Performance Between Ritual and Strategy", *Sociological Theory* **22** (4): 527–73.

Matthew Norton - 9781789906851

Alexander, Jeffrey C. and Philip Smith (1993), "The Discourse of American Civil Society: A New Proposal for Cultural Studies", *Theory and Society* **22** (2): 151.
Bearman, Peter (2012), "On Analytical Sociology", *Sociologica* **1**: 1–3.
Camic, Charles (1986), "The Matter of Habit", *American Journal of Sociology* **91** (5): 1039–87.
Dewey, John (1958), *Experience and Nature*. Courier Corporation.
Eliasoph, Nina and Paul Lichterman (2003), "Culture in Interaction", *American Journal of Sociology* **108** (4): 735–94.
Erikson, Emily (2017), "Networks and Network Theory: Possible Directions for Unification" in C.E. Benzecry, M. Krause and I.A. Reed (eds), *Social Theory Now*, Chicago: University of Chicago Press.
Fine, Gary Alan (1991), "On the Macrofoundations of Microsociology: Constraint and the Exterior Reality of Structure", *The Sociological Quarterly* **32** (2): 161–77.
Gibson, David R. (2011), "Avoiding Catastrophe: The Interactional Production of Possibility during the Cuban Missile Crisis", *American Journal of Sociology* **117** (2): 361–419.
Gross, Neil (2009), "A Pragmatist Theory of Social Mechanisms", *American Sociological Review* **74** (3): 358.
Gross, Neil (2013), "The Mechanistas", *Contemporary Sociology: A Journal of Reviews* **42** (3): 368–70.
Gross, Neil (2018), "The Structure of Causal Chains", *Sociological Theory* **36** (4): 343–67.
Hedström, Peter (2005), *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.
Hedström, Peter (2009), "The Analytical Turn in Sociology", in P. Hedström and B. Wittrock (eds), *Frontiers of Sociology*, Leiden: BRILL, pp. 331–42.
Hedström, Peter and Peter Bearman (eds) (2009), *The Oxford Handbook of Analytical Sociology*. New York: Oxford University Press.
Hedström, Peter and Petri Ylikoski (2010), "Causal Mechanisms in the Social Sciences", *Annual Review of Sociology* **36**: 49–67.
Hirschman, Daniel and Isaac Ariail Reed (2014), "Formation Stories and Causality in Sociology", *Sociological Theory* **32** (4): 259–82.
Joas, Hans (1996), *The Creativity of Action*. Chicago: University of Chicago Press.
Joas, Hans and Jens Beckert (2002), "Action Theory", in J.H. Turner (ed.), *Handbook of Sociological Theory*, *Handbooks of Sociology and Social Research*, Boston: Springer, pp. 269–85.
Kiser, Edgar (1994), "Markets and Hierarchies in Early Modern Tax Systems: A Principal-Agent Analysis", *Politics & Society* **22** (3): 284–315.
Kiser, Edgar and Michael Hechter (1998), "The Debate on Historical Sociology: Rational Choice Theory and its Critics", *American Journal of Sociology* **104** (3): 785–816.
Knight, Carly R. and Isaac Ariail Reed (2019), "Meaning and Modularity: The Multivalence of 'Mechanism' in Sociological Explanation", *Sociological Theory* **37** (3): 234–56.
Little, Daniel (2007), "Levels of the Social", in S.P. Turner and M.W. Risjord (eds), *Philosophy of Anthropology and Sociology*, *Handbook of the Philosophy of Science*, Amsterdam: Elsevier, pp. 343–71.
Little, Daniel (2008), "What Is 'Methodological Localism'?", *Understanding Society*. Retrieved August 28, 2020, https://undsoc.org/2008/11/02/what-is-methodological-localism/.
Little, Daniel (2012), "Analytical Sociology and the Rest of Sociology", *Sociologica* **1**: 1–47.
Little, Daniel (2014), "Actor-Centered Sociology and the New Pragmatism", in J. Zahle and F. Collin (eds), *Rethinking the Individualism-Holism Debate: Essays in the Philosophy of Social Science*, *Synthese Library*, Heidelberg: Springer, pp. 55–75.
Lizardo, Omar (2012), "Analytical Sociology's Superfluous Revolution", *Sociologica* **1**: 1–12.
Lizardo, Omar (2017), "Improving Cultural Analysis: Considering Personal Culture in its Declarative and Nondeclarative Modes", *American Sociological Review* **82** (1): 88–115.
Mahoney, James (2012), "The Logic of Process Tracing Tests in the Social Sciences", *Sociological Methods & Research* **41** (4): 570–97.
Manzo, Gianluca (2010), "Analytical Sociology and its Critics", *European Journal of Sociology* **51** (1): 129–70.
Manzo, Gianluca (2014), "Data, Generative Models, and Mechanisms: More on the Principles of Analytical Sociology", *Analytical Sociology*. Wiley, pp. 1–52.
Manzo, Gianluca (2020), "Agent-Based Models and Methodological Individualism: Are They Fundamentally Linked?" *L'Annee sociologique* **70** (1): 197–229.
McDonnell, Terence E. (2010), "Cultural Objects as Objects: Materiality, Urban Space, and the Interpretation of AIDS Campaigns in Accra, Ghana", *American Journal of Sociology* **115** (6): 1800–52.
Mead, George Herbert (1934), *Mind, Self & Society: From the Standpoint of a Social Behaviorist*, ed. C.W. Morris. Chicago: University of Chicago Press.
Norton, Matthew (2011), "A Structural Hermeneutics of 'The O'Reilly Factor'", *Theory and Society* **40** (3): 315–46.
Norton, Matthew (2014), "Mechanisms and Meaning Structures", *Sociological Theory* **32** (2): 162–87.
Norton, Matthew (2019), "Meaning on the Move: Synthesizing Cognitive and Systems Concepts of Culture", *American Journal of Cultural Sociology* **7** (1): 1–28.

Matthew Norton - 9781789906851

Peirce, Charles S. (1978), *The Collected Papers of Charles S. Peirce*, Vol. 5, ed. C. Hartshorne and P. Weiss. Cambridge: Harvard University Press.
Rawls, Anne Warfield (1987), "The Interaction Order Sui Generis: Goffman's Contribution to Social Theory", *Sociological Theory* **5** (2): 136–49.
Reed, Isaac A. (2013), "Power: Relational, Discursive, and Performative Dimensions", *Sociological Theory* **31** (3): 193–218.
Rochberg-Halton, Eugene (1986), *Meaning and Modernity: Social Theory in the Pragmatic Attitude*. Chicago: University of Chicago Press.
Rydgren, Jens (2009), "Beliefs", in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, New York: Oxford University Press.
Schneiderhan, Erik (2011), "Pragmatism and Empirical Sociology: The Case of Jane Addams and Hull-House, 1889–1895", *Theory and Society* **40** (6): 589.
Sewell, William H. (1996), "Historical Events as Transformations of Structures: Inventing Revolution at the Bastille", *Theory and Society* **25** (6): 841–81.
Sewell, William H. (2005), "A Theory of Structure: Duality, Agency, and Transformation", in *Logics of History: Social Theory and Social Transformation*. Chicago: University of Chicago Press, pp. 124–51.
Silver, Daniel (2011), "The Moodiness of Action", *Sociological Theory* **29** (3): 199–222.
Steinmetz, George (1998), "Critical Realism and Historical Sociology: A Review Article", *Comparative Studies in Society and History* **40** (1): 170–86.
Tavory, Iddo and Stefan Timmermans (2013), "A Pragmatist Approach to Causality in Ethnography", *American Journal of Sociology* **119** (3): 682–714.
Timmermans, Stefan and Iddo Tavory (2012), "Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis", *Sociological Theory* **30** (3): 167–86.
Vaughan, Diane (2004), "Theorizing Disaster: Analogy, Historical Ethnography, and the Challenger Accident", *Ethnography* **5** (3): 315–47.
Wagner-Pacifici, Robin (2010), "Theorizing the Restlessness of Events", *American Journal of Sociology* **115** (5): 1351–86.

Matthew Norton - 9781789906851

9. Action

Clemens Kroneberg and Andreas Tutic[1]

1. ACTION THEORY AND ANALYTICAL SOCIOLOGY

At the heart of analytical sociology are social mechanisms, of which actions are essential building blocks. It is therefore no wonder that definitions of analytical sociology usually emphasize the role of actions (Hedström and Swedberg 1998; Hedström and Bearman 2009: 4). In the well-known macro-micro-macro model, the explanation of actions constitutes the intermediate step on the micro level that is essential to open up the black box of macro-level relationships and to provide a deeper understanding of collective phenomena: Parties and citizens' behavior underlies electoral outcomes, residential, economic and political choices generate neighborhood change, and consumers and producers' choices make for fads and fashions.

Analytical sociology seeks precise, abstract and action-based explanations of social phenomena (Hedström and Swedberg 1998). The emphasis on precise and abstract models distinguishes it from other important sociological traditions: Compared to interpretive, phenomenological or pragmatist approaches, analytical sociology is much stronger oriented toward the sciences and less so toward the humanities. Rather than viewing social theory as mainly providing concepts, heuristics and frameworks, analytical sociology aims at generative models that specify how actions and interactions among interrelated entities (e.g., individuals or corporations) give rise to collective phenomena. This explanatory ambition means that analytical sociology is weakly concerned with questions, such as "What is Agency?" (Emirbayer and Mische 1998) or "What is Action?" (Joas and Knöbl 2009). Theoretical progress is not sought in increasingly differentiated, complex or nuanced conceptualizations of actions, practice or communication (Healy 2017). Rather, explanations of actions have to identify (proximate) causes, thereby answering why individuals behave in certain ways, rather than others. Accordingly, analytical approaches in sociology—as well as in economics and political science—usually view actions as alternatives among which a selection takes place.

Analytical sociology shares these meta-theoretical principles—the need for microfoundations, as illustrated in the macro-micro-macro model, precision and abstraction—with its older sibling: Rational Choice Theory (RCT). RCT is the most elaborated action-theoretic paradigm in the social sciences. While some scholars have found it difficult to tell RCT and analytical sociology apart (Opp 2013), a major difference actually lies in the role assigned to theories of action. As this question is somewhat controversial even within analytical sociology, it is helpful to distinguish a weak and a strong program of action theory in analytical sociology.

The weak program states that, although analytical sociology "emphasizes the importance of action in the explanation of social phenomena, it does not subscribe to an axiomatic vision according to which a specific action theory should be used for all purposes" (Hedström and

[1] Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany 's Excellence Strategy—EXC 2126/1–390838866.

Clemens Kroneberg and Andreas Tutic - 9781789906851

Ylikoski 2014: 64). Hence, despite the emphasis on actions as building blocks of social mechanisms, this variant of analytical sociology is characterized by a strategic disinterest in developing action theory. Its strategic nature stems from two advantages: First, it allows analytical sociology to escape the largely fruitless debates about "rationality" that have occupied rational choice scholars, especially within sociology. Second, it involves a shift in emphasis within the macro-micro-macro model: Rather than zooming in on the explanation of individual behavior, a more pragmatic approach to this step allows one to direct even more attention and effort to examining the social dynamics in which the actions of embedded individuals give rise to macro-level consequences.[2]

To further illustrate this, consider two main types of agent-based models employed in analytical sociology: One class of abstract simulation models aims to identify conditions that are sufficient to generate particular collective outcomes (sociological methodology). In these models, simple rules are assumed to govern the behavior of agents. For example, in Schelling's classic segregation model, agents are assumed to move out of their neighborhood once the fraction of outgroup members exceeds their individual threshold (Schelling 1971). And in Axelrod's model of cultural differentiation, agents are assumed to be more likely to interact with neighbors the more culturally similar they are and then to assimilate to one of their cultural traits due to social influence (Axelrod 1997). Although one could in principle use elaborated theories of action to derive such rules, these models realize their explanatory aims without doing so. Another class of models are empirically calibrated action-based models (ABMs) that use empirical data to ground the parameters of action rules or even the rules themselves (see Bruch and Atwell 2015). For example, Bruch and Mare (2006) build an empirically calibrated version of Schelling's segregation model by using vignette data from surveys to estimate how black and white respondents' neighborhood preferences respond to the percentage of outgroup members. In either class of models, the simplicity and realism of behavioral rules is all that matters. As analytical sociology does not subscribe to a model of scientific explanations that requires the application of laws (i.e. the covering-law or Hempel–Oppenheim model), there is no need to derive such rules from law-like generalizations about human behavior (see Kalter and Kroneberg 2014).[3]

The strong program of action theory in analytical sociology also departs from RCT. However, it holds onto the desirability of a theory of action that allows for differentiated explanations of human behavior and to approach a great variety of substantive fields with a unified action-theoretic core (Esser 2009; Lindenberg 2013; Wikström 2014). In the search for realistic microfoundations, proponents of a strong program of action theory join other disciplines

---

[2] This emphasis on parsimony is shared by major advocates of rational choice theory. As Coleman (1990: 19) put it: "For a social theory made up of three components—a macro-to-micro component, an individual-action component and a micro-to-macro component—it is especially important that the individual-action component remain simple. […] I have chosen to trade off as much psychological complexity as possible in order to allow introduction of greater amounts of complexity in the other two components of the theory, the "social organizational" components." However, Coleman still held on to the desirability of a unifying theory of (rational) action.

[3] Watts (2014) has put forward a more fundamental criticism of action theories in sociology. Rather than challenging the agenda of action theories per se, his criticism highlights a number of methodological imperatives: Most importantly, collective outcomes should not be attributed to the proximate causes of individual behavior (as this would ignore the micro-macro transition), assumptions about the relevant mental constructs should be grounded in empirical data or more specific theoretical knowledge (on so-called bridge assumptions, see Kroneberg and Kalter 2012), and rigorous tests of action theories should involve specific predictions tested on new data (rather than mere ex-post accounts).

Clemens Kroneberg and Andreas Tutic - 9781789906851

such as cognitive psychology or behavioral economics (see Manzo 2013). Proponents of this research line do not deny that the study of complex social dynamics requires using relatively simple models of individual behavior (Lindenberg 1992; Kroneberg 2014). However, to avoid ad hocism and a proliferation of unrelated behavioral rules, such models should be derived from more general theories that allow researchers to introduce simplifying assumptions in a controlled and transparent manner. This kind of modularity allows for flexible adaption to explanatory needs and emphases in the macro-micro-macro model (see Kroneberg 2014).

The remainder of this chapter is organized as follows: After sketching a fundamental challenge facing analytical theories of action, we summarize how RCT explains human behavior. Against this background, we describe recent attempts to formulate more realistic theories of actions in the strong program of action theory within analytical sociology and related approaches.

2. THE FUNDAMENTAL CHALLENGE OF ANALYTICAL ACTION THEORIES

For a scientific observer, the problem of explaining human action presents itself as depicted in Figure 9.1. That is, we observe an actor taking a course of action against the background of a choice situation. Action theorists explain the observable relationship between choice situations and actions by postulating certain mental constructs as well as a system of rules which specifies actions for all conceivable choice situations and all configurations of mental constructs. Hence, the fundamental idea of action theory is to conceptualize actions as depending not only on directly observable properties of the choice situation, but also on not directly observable internal states of the actor. Clearly, any empirical science has to take extra methodological care in dealing with unobservable concepts such as mental constructs.

Sociology is rich in mental constructs; a great many of these such as values, attitudes, norms, emotions, habits or memories have been proposed to account for human behavior. At the

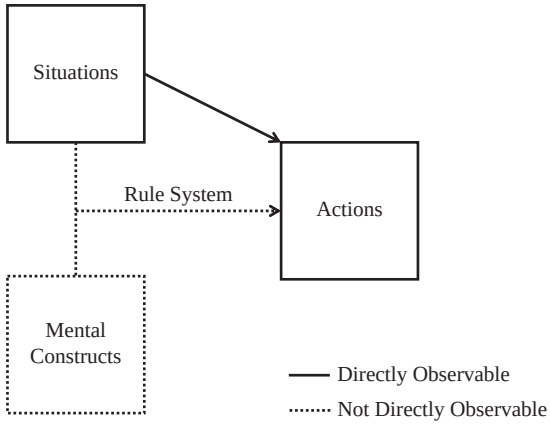


Figure 9.1 The logic of analytical action theory

Clemens Kroneberg and Andreas Tutic - 9781789906851

same time, sociology is relatively poor in rule systems, which specify the interplay of mental constructs and properties of choice situations in determining behavior. Most action theories in sociology are content in providing a set of mental constructs that, in one way or the other, "impact" or "influence" action. This is especially true of those sociological paradigms that focus on understanding human action (Weberian *verstehen*). For analytical sociology with its aspiration to causally explain human action (Weberian *erklären*) as well as social phenomena, a proper action theory must necessarily include an explicit rule system. Formulating such rule systems and devising methodologies to apply and test them are core tasks of analytical action theories.

3. RATIONAL CHOICE THEORY

RCT has most fully lived up to the task for developing and refining analytical models of action. As is widely acknowledged there are contending conceptions of RCT, including controversies about the substantive meaning and methodological status of its core concepts such as "rationality", "utility" and "subjective probabilities" (Goldthorpe 1998; Opp 1999; Boudon 2003). In our subsequent exposition, we will adopt a modern theory-constructive point of view. According to this view, RCT consists of a set of interrelated mathematical theories, such as decision and game theory, which serve as a *toolbox* for constructing applied theories in the social sciences (and beyond, for instance in biology). An important consequence of this perspective is that RCT as a mathematical theory cannot be refuted by empirical observations. All that can be tested in empirical studies are applications of RCT to social phenomena. These applications model certain social phenomena as mathematical objects. For example, human beings are interpreted as carriers of preference relations and subjective beliefs, social situations are interpreted as games, and equilibrium concepts are interpreted as behavioral predictions. In the following, we limit our exposition to basic instruments of RCT in modelling individual behavior.

3.1 Rationality and the Preference-Constraints Model (PCM)

The defining core feature of modern RCT is that behavior is modelled as the maximization of a preference relation under constraints (cf. Kreps 1988; Rubinstein 1998, 2017).

Definition 1. Let X be a nonempty set. A binary relation $\succsim \subseteq X \times X$ *is called a weak preference on X, if the following two conditions are satisfied:*

- *Completeness*: For all $x, y \in X$ it holds: $x \succsim y$ or $y \succsim x$.
- *Transitivity*: For all $x, y, z \in X$ it holds: $x \succsim y$ and $y \succsim z$ implies $x \succsim z$.

Given a weak preference $\succsim$, we can define the associated strict preference $\succ$ by letting $x \succ y$ if and only if $x \succsim y$ and not $y \succsim x$. Similarly, the associated indifference relation $\sim$ is defined by $x \sim y$ if and only if $x \succsim y$ and $y \succsim x$. Note that $\succsim = \succ \cup \sim$ and $\succ \cap \sim = \emptyset$.

Objective situations are simply modelled as subsets of X. The notion of constrains indicates that in a decision situation $S \subseteq X$, the decision-maker is constrained in his/her available choices, i.e. s/he cannot choose an element from $X \backslash S$.

Clemens Kroneberg and Andreas Tutic - 9781789906851

Definition 2. Let X be a nonempty set. The set of decision problems on X is defined by $\boldsymbol{X} := \boldsymbol{2}^X \setminus \{\emptyset\}$.

Given this setup of mental construct (preference relation) and objective situation (decision problem), RCT posits a very basic and intuitive decision rule:

Definition 3. Given a preference $\succsim$ *on X and a decision problem* $S \in \boldsymbol{X}$, *the set of rational choices in S is defined by* $C^{\succsim}(\mathcal{S}) := \{s \in \mathcal{S} : \nexists s' \in \mathcal{S} \text{ s.t. } s' \succ s\}$. *This definition induces a rational choice function* $C^{\succsim} : \boldsymbol{X} \to 2^X$ *satisfying* $C^{\succsim}(\mathcal{S}) \subseteq \mathcal{S}$ *for all* $S \in X$.[4]

Any behavior which can be described as the maximization of a preference relation under constraints is defined as *rational behavior* (Rubinstein 1998). Pause for a second and let that sink in: In the modern perspective on RCT, rationality means nothing more than *describability* by this very basic calculus. Rationality does not imply any specific "content" of the preferences such as material egoism. Altruism and orientation toward attitudes, norms and values are rational as long as the specifics of these orientations do not undermine the describability of the resulting behavior by PCM. Moreover, rationality does not even imply any form of deliberate, meaningful, reasoned, or "instrumentally rational" action (Gintis 2017). Hence, RCT is not a psychological theory of cognition.

Still, the application of PCM to human behavior comes with considerable empirical implications; that is, not any pattern of behavior can be modelled this way. The calculus can reproduce only patterns of behavior which satisfy certain consistency requirements. To see this more clearly, we need to formalize the notion of patterns of behavior first:

Definition 4. A function $C: \boldsymbol{X} \to \boldsymbol{X}$, *which satisfies* $C(S) \subseteq S$ *for all* $S \in \boldsymbol{X}$, *is called a decision function on X.*

Clearly, not every decision function can be reconciled with PCM. This raises the following question: What properties does a decision function need to exhibit such that there exists a preference relation which can explain this decision function? From a methodological point of view, identifying these conditions is highly desirable, since they pinpoint the empirical scope of RCT. Such questions are the principal subject of *axiomatic decision theory* and the *Theory of Revealed Preferences* (cf. Kreps 1988), which form the analytical core of RCT. With respect to the basic PCM, the respective proposition was established by Samuelson (1938):

Proposition 1. Let C be a decision function on X. If and only if C satisfies the Weak Axiom of Revealed Preference (WARP), there exists a preference $\succsim$ *on X such that* $C^{\succsim} = C$.

Definition 5. Let C be a decision function on X. C satisfies ***WARP****, if for* $x,y \in X$ *it holds: If there exists some* $S \in \boldsymbol{X}$ *such that* $x, y \in S$, $x \in C(S)$ *and* $y \notin C(S)$, *then there does not exist some* $T \in \boldsymbol{X}$ such that $x, y \in T$ and $y \in C(T)$.

---

[4] Note that $C^{\succsim}(\cdot)$ might be a non-singleton (due to indifferences) and also be empty. For instance, if we take $\geq$ as a weak preference on $\mathbb{R}$, we find that $C^{\succsim}([0,1)) = \emptyset$, because for any $s \in [0,1)$ there is some s' satisfying $s < s' < 1$. Both non-uniqueness of rational choice and its non-existence can pose considerable difficulties in applications of this calculus to social phenomena.

Clemens Kroneberg and Andreas Tutic - 9781789906851

Note that, crucially, **WARP** is not a property of mental constructs (preference relations), but a property of an empirically observable phenomenon (behavior). **WARP** identifies precisely what rationality means on a purely behavioral level, thereby also setting a normative standard as to what kind of behavior is considered rational by RCT.

Besides identifying the empirical scope of PCM, Proposition 1 or, to be more precise, its proof satisfies a second major methodological desideratum: It provides a measurement theory for the postulated mental constructs (in this case for a preference relation). Note that Proposition 1 implicates that if C satisfies **WARP** there exists some $\succsim$ such that $C^{\succsim} = C$. The only way to prove this is to construct (i.e. "to measure") a relation R purely on the basis of C, demonstrate that R satisfies the formal requirements of a preference and finally show the equality $C^R = C$. In this case, the measurement is straightforward: For all xRy simply define xRy if $x \in C(\{x,y\})$. In other cases, the measurement procedures are much more involved.

3.2 Utilities and Probabilities

Under certain technical conditions, which are studied in ordinal utility theory (Ok 2007), preference relations can be represented by numerical functions called utility functions.

Definition 6. Let $\succsim$ be a preference on X. A function u: $X \to \mathbb{R}$ is called a utility function for $\succsim$, if for all $x, y \in X$ it holds: $x \succsim y$ if and only if $u(x) \geq u(y)$.

In principle, utility functions are merely a convenient way of representing assumptions about preferences in applications and, under certain conditions, allow the application of calculus, which simplifies finding solutions for problems of constrained maximization. That is, utility functions are tools that simplify the application of PCM and, more generally, RCT to social phenomena. Conceptually, the notion of utility adds nothing to the underlying action theory.

Note that historically the notion of "utility" precedes the modern concept of preferences (Read 2007). In fact, PCM was introduced to lay a solid foundation for the notion of utility and clarify certain misconceptions of this term in older writings. Sparing the details of outdated discussions, let us simply note that any inference drawn from the utilities of actors needs to take into account the scale level on which utilities are measured. Utility functions for PCM set-up are unique up to positive monotone transformations (ordinal scale level).

Cardinal utility theory is the final major building block of decision theory and deals with the problem of modelling decision-making under risk and uncertainty. In PCM the actor has control over all that interests her, i.e. s/he chooses among alternatives over which his/her preference relation is defined. In many applications, this setup is too restrictive. An agent might be interested in consequences of his/her actions, but the consequences might only partially be determined by his/her actions.

Definition 7. Let X be a finite and nonempty set of consequences. A lottery is a discrete probability distribution on X, i.e. a function l: $X \to [0,1]$ with $\Sigma_{x \in X} l(x) = 1$. The set of all lotteries on X is denoted by L.[5]

[5] Note that L can also be interpreted as a subset of $\mathbb{R}^{|x|}$, i.e. a lottery can also be interpreted as a vector; hence we can perform the usual algebraic manipulations on lotteries.

Suppose we have a preference on X and are interested in constructing a preference on L on this basis. Cardinal utility theory takes a quite intuitive approach by describing a distribution by its expectation. That is, the main results in cardinal utility theory clarify the conditions under which choice behavior among lotteries can be represented as the maximization of an expected utility function, which assigns utilities to risky or uncertain lotteries by taking probability-weighted averages of utilities of consequences.

Proposition 2. A continuous[6] *preference* $\succsim$ *on L satisfies* ***IIA*** *if and only if there is a function* $u: X \rightarrow \mathbb{R}$ *such that for all* $p, q \in L$ *it holds:* $p \succsim q \Leftrightarrow \Sigma_{x \in X} p(x)u(x) \geq \Sigma_{x \in X} q(x)u(x)$.

Note that this proposition establishes conditions under which lotteries on consequences are "evaluated" by their expected means measured in utilities. Clearly, the utilities in this representation must be measured on a higher scale level than the utilities for PCM. That is, $u: X \rightarrow \mathbb{R}$ is measured on an interval scale, meaning that utilities are unique up to positive affine transformations. The proof of Proposition 2 includes a measurement theory for u, which can also be implemented in the lab and, in principle, even in surveys.

As should be expected, the consistency requirements imposed by such a representation to be viable are much more involved than those requirements needed to justify ordinary rational choice behavior. Using the expected utility calculus in applications means that the social scientist assumes that choice among risky prospects is not only consistent in the sense of **WARP**, but also consistent in the sense of a property called Independence of Irrelevant Alternatives.[7]

Definition 8. A preference $\succsim$ *on L satisfies Independence of Irrelevant Alternatives (**IIA**), if for all* $p, q, r \in L$ *and any* $\alpha \in (0,1]$ *it holds:* $p \succsim q \Leftrightarrow \alpha p + (1-\alpha) r \succsim \alpha q + (1-\alpha) r$.

IIA viciously enforces the applicability of modelling choice behavior via the linear-additive expected utility calculus. The complex lotteries $\alpha p + (1 - \alpha)r$ and $\alpha q + (1 - \alpha)r$ are to be evaluated solely on the basis of their differences (p and q), while their common component (r) has to be ignored.

So far our treatment of cardinal utility theory worked with objective probabilities, which can be seen from the fact that in the setup described by Definitions 7 and 8 as well as by Proposition 2 the lotteries on X are exogenously given. *Subjective Expected Utility Theory* extends the expected utility representation to situations in which subjective probabilities are involved (Savage 1954). Besides consistency requirements in the line of **WARP** and **IIA**, which are needed to found the utilities, there are very strong additional consistency requirements, such as the famous *Sure Thing Principle*, to justify the notion of subjective probabilities.

3.3 Sociological SEU Theory

In sociological RCT it has been quite popular to rely on some kind of expected utility calculus in representing determinants of actions, such as educational choices (Breen and Goldthorpe

[6] Speaking loosely, a preference is continuous if it ranks similar objects similarly. Of course, this notion presumes some kind of metric on X, which is absent in the basic PCM. L as a subset of $\mathbb{R}^{|X|}$ has a metric.

[7] Assuming continuous preference relations is not as innocent as it might seem; for instance, lexicographic preferences, which are sometimes advocated to model normative orientations (Voss 1985), fail continuity. Still, **IIA** is commonly considered the more problematic consistency requirement.

1997), political participation (Finkel et al. 1989) or migration (Esser 2004). Scholars often describe the underlying action theory as SEU theory. However, this parlance is somewhat misleading since this kind of SEU theory differs considerably from the mathematically expected utility theory.

The setup of sociological SEU theory seems to be as follows: There is a set of actions A and a set of consequences X. The decision-maker carries an utility function $u: X \rightarrow \mathbb{R}$. In addition, for each action $a \in A$, there exists a "probability function" $p_a: X \rightarrow [0,1]$, which gives the subjective probability held by the decision-maker as to how likely the realization of each consequence is, given s/he take action a. We are putting probability function in quotation marks here because $\Sigma_{x \in X} p(x) = 1$ is generally not required; it seems that an element $x \in X$ is typically not interpreted as an all-encompassing state of the world, but as an aspect of such states. Sociological SEU theory assumes or hypothesizes that a decision-maker chooses the alternative $a \in A$, which maximizes the so-called SEU weight $SEU(a) = \Sigma_{x \in X} p_a(x) u(x)$.

Sociological SEU theory is much closer to value-expectancy theories in social psychology, e.g. the *Theory of Planned Behavior* (Ajzen 1991), than to mathematically expected utility theory. In particular, there is no representation theorem along the lines of Proposition 2 for sociological SEU theory. From the perspective of mathematical decision theory, this implies a lack of rigor and testability as it remains unclear what kind of patterns of behavior can (or cannot) be modelled by this formalism and how the involved mental constructs can be measured. Hence, "the 'wide' sociological variants of rational choice theory have failed to develop a methodological framework of comparable analytical power" (Manzo 2013: 370).

In empirical research, sociological SEU theory is often tested by measuring both utilities and subjective "probabilities" via ordinal Likert scales in surveys and using regression models to examine their associations with self-reported behaviors (the so-called direct test strategy, see Kroneberg and Kalter 2012). To the extent that the coefficients of benefits, costs and beliefs are statistically significant and exhibit the hypothesized signs, such statistical analyses are seen as supporting SEU theory. Sometimes it is also tested whether benefits and costs interact with beliefs so that the former are more relevant the higher their subjective probability. And in many sociological applications, a crucial question is whether including such direct measures accounts for the effects of background factors, such as social origin effects on educational choices or gender effects on fields of study. Hence, sociological SEU theory offers explicit microfoundations and ways to assess its empirical significance—even though the standards of evidence (average associations between variables) are very different and less rigorous than those employed in decision theory.

4. SOCIOLOGICALLY AND PSYCHOLOGICALLY RICHER MICROFOUNDATIONS

Despite its fruitfulness in applied research and its theoretical coherence, recent decades have seen considerable efforts in transcending RCT and in particular decision theory. After briefly describing the main reasons for these aspirations, we will outline two major research programs along these lines, which appear to be highly compatible with analytical sociology: models of Bounded Rationality and the Dual-Process Perspective.

Clemens Kroneberg and Andreas Tutic - 9781789906851

4.1 Shortcomings of Orthodox Decision Theory and Sociological Rational Choice Theory

Empirical and experimental research in particular have demonstrated that human cognition and behavior systematically deviate from the consistency requirements of standard formalisms. With respect to decision theory, canonical and well-replicated anomalies such as the *Allais Paradox* or the *Ellsberg Paradox* provide direct empirical tests of particular consistency requirements such as **IIA** and the Sure Thing Principle and show that a considerable amount and often the majority of subjects violate these properties. By now, convincing anomalies are documented in the literature for basically all of the consistency requirements of orthodox decision theory (for example, Sippel 1997; Rubinstein 2013). In addition, RCT makes use of other formalisms such as logic and stochastics and these formalisms are *built into*, i.e. deeply integrated in, the RCT formalism. To just give one example, the *sequential equilibrium* in game theory assumes that players update beliefs according to Bayes's rule. However, the *Heuristics and Biases Program* initiated by Amos Tversky and Daniel Kahneman has clearly demonstrated that, even in simple problems, people make systematic errors, e.g. in updating beliefs and drawing logically correct inferences. These findings demonstrate that the calculi employed by proponents of RCT such as the expected utility principle, logic and statistics fail to grasp the empirical diversity in human behavior and cognition. Within the field of decision theory, such anomalies have motivated a family of models that set out to represent humans' Bounded Rationality.

As described, sociological RCT constitutes a style of theory-guided social research that is quite different from experimental lab studies tailored to test axiomatic decision theories. However, in this context, too, criticism has been voiced against explaining behavior solely with reference to actors' preferences, beliefs and constraints. Most importantly, although this conceptual parsimony has important advantages in the macro-micro-macro model, it has little heuristic or hypotheses-generating power and tends to yield trivial explanations when the explanatory focus is mainly on individual behavior. A famous example is the explanation of electoral participation as reflecting the costs and expected utility of participation, which includes a preference to act in accordance with one's civic duty. Concepts such as values, attitudes, norms, emotions, habits or memories are not primitive concepts of RCT. As a consequence, RCT can only integrate these concepts and their impact on behavior via modelling them as part of preferences and subjective beliefs. While advocates of RCT have been keen to highlight the openness of preferences and subjective beliefs to incorporate mental constructs described in alternative paradigms (Opp 1999), it has to be kept in mind that RCT per se offers *no systematic guideline* of doing so. Invoking a civic duty norm and extending the utility calculus by psychological incentives has therefore been met with criticism within and outside of RCT. In addition, it has been argued that the workings of normative commitments and other mental constructs cannot be fully captured by the formalism of RCT (Elster 1989). While norms can be modelled as part of the utility function of actors, this implies that the actor is willing to make trade-offs between norm conformity and other considerations such as material consequences (Kroneberg et al. 2010a: 6). However, this conception is at odds with the idea of unconditional norm compliance, which is central to many action-theoretic accounts in sociology (e.g. *Wertrationalität*; Weber [1922] 1972; Parsons 1937). Hence, the parsimony of RCT means that it yields few hypotheses on the interplay of different determinants of action, such as cultural schemas, moral norms, emotions or expected consequences. Additionally, this limits its capability to integrate concepts and insights from alternative paradigms in the social sciences.

Clemens Kroneberg and Andreas Tutic - 9781789906851

In sum, both orthodox decision theory as well as sociological RCT suffer from significant shortcomings. We now turn to two families of models that have taken these limitations as starting points to develop more adequate and richer theories of action.

4.2 Bounded Rationality

Under the impression of the great variety of anomalies of RCT uncovered by the Heuristics and Biases Program, the last two decades have seen a renewed interest in alternative decision theories, grouped under the label *Bounded Rationality*. While the older literature on Bounded Rationality was quite contempt in a pure description of alternative decision procedures, modern accounts in Bounded Rationality actually aim at axiomatization (Rubinstein 1998). This new literature on axiomatic theories of Bounded Rationality provides concrete examples as well as explicit and implicit guidelines as to how alternative action theories with richer assumptions regarding the mental life of actors can be formalized while preserving key methodological advantages of orthodox decision theory.

In technical terms, many of these alternative decision procedures are extensions of the classic PCM: They posit additional mental constructs, beyond preference relations (and subjective probabilities), coupled with decision rules that are compatible with classical rationality but also with different logics of decision-making. To make matters more concrete, we consider the example of PCM with Limited Attention (Masatlioglu et al. 2012; Lleras et al. 2017). This decision procedure provides a basic model of a decision-maker who is not necessarily aware of all available choice alternatives.

Definition 9. A decision-maker exhibiting PCM with Limited Attention is characterized by a pair ($\succ$, Γ), in which $\succ$ is a quasi-complete[8] *strict preference on X and Γ is a function Γ: $\boldsymbol{X} \rightarrow \boldsymbol{X}$ such that $\emptyset \neq \Gamma(S) \subseteq S$ for all $S \in \boldsymbol{X}$. Γ is required to satisfy the following condition (Irrelevance; IR): $\forall S \in X: s \notin \Gamma(S) \Rightarrow \Gamma(S \setminus \{s\}) = \Gamma(S)$. PCM with Limited Attention induces a decision function by letting $C^{\succ,\Gamma}(S) := C^{\succ}(\Gamma(S))$ for all $S \in \boldsymbol{X}$.*

The filter Γ is used to model the attention of the decision-maker. In choice problem S s/he subjectively take only the options in $\Gamma(S)$ into consideration, although objectively the alternatives in $S \setminus \Gamma(S)$ are also available. There might be multiple underlying reasons for the phenomenon of limited attention; for instance, limited attention might be due to an underinvestment in the gathering of information due to urgency, i.e. an emotionally grounded preference for earlier action over later action (Elster 2009: 68). Faced with a decision situation, the decision-maker picks the most preferred alternative among those s/he is aware of. Note that the filter Γ is an additional mental construct that this alternative decision procedure attributes to decision-makers. Condition *IR* is an assumption regarding the mental life of decision-makers, which can in principle be judged against the background of psychological research. Prima facie, *IR* is quite appealing: If the decision-maker is not aware of some objective alternative in a choice problem, its removal should not alternate the set of alternatives the decision-maker is aware of.

In perfect analogy to the traditional Theory of Revealed Preferences, Masatlioglu et al. (2012) provide a characterization of PCM with Limited Attention in terms of overt behavior.

---

[8] A binary relation R on X is quasi-complete if for all $x,y \in X$ with $x \neq y$ it holds: xRy or yRx. Essentially, PCM with Limited Attention abstracts from indifferences. Hence, some form of completeness of the strict preference is needed.

It turns out that any behavioral pattern that is characterized by the following property can be explained by this particular decision procedure:

*Definition 10. A choice function C satisfies WARP with Limited Attention (**WLA**) if for all* $S \in X$ *there exists some* $s \in S$ *such that for all* $T \in \boldsymbol{X}$, $|T| > 1$ *with* $s \in T$: $C(T) \in S$ *and* $C(T \setminus \{s\}) \neq C(T) \Rightarrow C(T) = s$.[9]

In formal terms, **WLA** is a weakening of **WARP**, i.e. any choice function that can be explained by PCM can also be explained by PCM with Limited Attention, but not conversely. Note also that because of this characterization result by Masatlioglu et al. (2012) PCM with Limited Attention shares a core methodological advantage with traditional PCM: We know exactly which kind of behavior is explainable by this alternative decision procedure and which kind of observations would contradict this model. In addition, the proof of their characterization involves a measurement theory for preferences and *filters*; given a list of observed behaviors, we can make theoretically solid inferences regarding these postulated mental constructs.

PCM with Limited Attention is but only one example of many interesting alternative decision procedures in the axiomatic theories of Bounded Rationality. Other procedures model the influence of default options (Rubinstein and Zhou 1999; Masatlioglu and Ok 2005), decision-making under the influence of "frames" (Salant and Rubinstein 2008) as well as instinctive decision-making (Sandroni 2011), just to name a few. There are also well-founded alternatives to (subjective) expected utility theory, such as axiomatizations of different versions of Kahneman and Tversky's Prospect Theory (see Wakker 2010).

For analytical sociology, which emphasizes precision and abstraction while being skeptical about the realism of traditional RCT, theories of Bounded Rationality provide a substantial enrichment in techniques to conceptualize and model individual decision-making. In contrast to other alternatives to orthodox RCT, such as sociological SEU theory, axiomatic theories of Bounded Rationality rest on a solid mathematical foundation and have all the methodological merits of traditional decision theory. Most importantly, they show how the methodology of the Theory of Revealed Preferences can be used in formalizing alternative and descriptively more accurate decision procedures.

4.3 Dual-Process Perspective

While theories of Bounded Rationality provide mathematically well-founded alternative decision procedures resting on psychologically informed richer microfoundations, they provide little guidance on how to integrate this plethora of decision procedures into an integrated and unified model of human behavior. In addition, there is *Behavioral Economics*, which, in trying to conserve as much as possible from orthodox RCT while coming to terms with "anomalous" empirical findings, offers a great variety of specific assumptions about social preferences (e.g. Bolton and Ockenfels 2000; Fehr and Schmidt 1999).

This proliferation of specific "decision theories" seems to underscore doubts about the feasibility of the project of a general theory of action (e.g. Hedström and Ylikoski 2014).

[9] Note that, for simplicity, we are abusing notation here: Originally, we defined C as mapping to $\boldsymbol{X}$, but in this context it maps to X. This is common practice when dealing with decision procedures that always output a single action alternative.

Clemens Kroneberg and Andreas Tutic - 9781789906851

However, there is also reason to advocate the idea that a promising synthesis is actually in the making: The *Dual-Process Perspective* (DPP) on human cognition and behavior. Our advocacy of DPP stems from the fact that there is a convergence of three literatures towards an integrative theory of action: The Dual-Process approach in cognitive and social psychology (Evans 2010; Kahneman 2011; Stanovich 2011), recent theories of action in analytical sociology (Esser and Kroneberg 2015), and many decision procedures in axiomatic theories of Bounded Rationality (Rubinstein 1998, 2013; Sandroni 2011) rest on the idea of explaining human cognition and human behavior from the interaction of two *modes* or two "*systems*" or two kinds of *processes*. System 1 consists of autonomous, fast and usually associative processes of cognition, which take place predominantly outside the consciousness of the actor. System 2 comprises cognitive processes that are under the control of the actor, slow and consist in the conscious manipulation of certain calculi and formalism inside of the working memory of the actor. Note that while automatic Type 1 processes of cognition and behaving lack rationality in the sense of deliberate reflection, in many instances they might exhibit "ecological rationality", rationality in the sense of being well adapted to the environment of the actor and providing him "fast and frugal" means of dealing with it (Gigerenzer and Todd 1999; Kenrick and Griskevicius 2013).[10]

In analytical sociology, the *Model of Frame Selection* (MFS), introduced by Hartmut Esser (1991, 1996), is the most elaborated version of DPP (see Kroneberg 2005, 2014). In its full version, the MFS explains behavior by a series of three selections (that of a frame, a script and an action), each of which can be processed in either the automatic-spontaneous mode (as-mode) or the reflective-calculating mode (rc-mode). In the frame selection, actors interpret and define the situation based on situational cues. Frames are mental models of typical situations both of biological and socio-cultural origin and differ in the extent to which they match a given objective situation. Based on an actor's definition of the situation, the selection of a script takes place. Scripts are mental models of typical sequences of actions such as biological instincts, routines or social norms. Selecting a script does not necessarily mean behaving in accordance with this script; it merely indicates that the script is (either consciously or unconsciously) present in the mind of the actor and thereby influences subsequent choice behavior. Importantly, the activation weight of a script depends on the selected frame, which is to say that the orientation towards action sequences depends on the definition of the situation. Once a script selection is achieved, an actual choice among behaviors is conducted.

Importantly, each of these selections can be processed in the as-mode (Type 1 process) or the rc-mode (Type 2 process). If the whole sequence takes place in the as-mode, the selection of a frame, a script and a corresponding action is a process of spreading activation. Actors

[10] Strictly speaking, Gigerenzer does not consider himself to be an advocate of DPP and instead develops an alternative theoretical paradigm which conceptualizes both intuitive and deliberate judgments as being rule-based (Kruglanski and Gigerenzer 2011). At the same time, scholars working within DPP argue that Gigerenzer tends to downplay qualitative differences between Type 1 and Type 2 processes that call for different calculi to model these processes (Evans and Stanovich 2013: 231). For the purpose of this overview, there is no need to engage with the intricacies of this debate within the psychological literature. Both the fast and frugal research program and DPP highlight the importance of heuristic reasoning and decision-making that depart from classic calculi, and both agree that the crucial task for theoretical as well as empirical research is to determine the conditions under which such intuitive forms of decision-making prevail.

Clemens Kroneberg and Andreas Tutic - 9781789906851

behave spontaneously without deliberating on competing action alternatives.[11] In contrast, the rc-mode allows one to represent processes of deliberative decision-making that are controlled, slow, based on calculi and conscious. Esser and Kroneberg (2015) suggest that many selections in the rc-mode can be modelled by traditional RCT and sociological SEU theory.

A particularly noteworthy feature of MFS is that it is quite explicit with respect to a critical theoretical problem of the DPP. Postulating two kinds of mental processes is one thing; providing explicit conditions under which each of these types of processes comes about is another. Kroneberg (2005) derives the following inequality from an application of sociological SEU theory to the problem under which conditions the selection of a frame occurs in the as-mode:

$$m \geq 1 - \frac{C}{p \cdot U}$$

in which m is the match of the most suitable frame, C measures the costs of reflection, p measures the opportunities for reflection and U measures the motivation for reflection. According to this inequality, a spontaneous definition of the situation is more likely, the greater the costs of reflection and the smaller the opportunities and motivation for reflection are.

This inequality also implies a hypothesis that corresponds with a canonical idea in sociological action theory: If the match of frame is close to 1, the definition of the situation is bound to be processed in the as-mode. When this strong spontaneous activation extends to a script that regulates the behavior of interest, actors will not consider competing action alternatives and incentives. Hence, the MFS allows for an unconditional effect of cultural orientations on behavior and specifies its scope conditions. For instance, Kroneberg et al. (2010a) demonstrate that citizens who have strongly internalized the civic duty norm tend to participate in elections irrespectively of the extent they identify with a party and irrespectively of their perceived influence on the outcome of the vote. The latter two variables, which are well-established predictors on voting behavior, only affect citizens with a relatively weak internalization of the civic duty norm. Similar findings, in which strong cultural orientations suppress the effects of incentives, have been uncovered in numerous empirical applications of MFS in areas such as helping behavior in high-cost situations, deviant and criminal behavior or fertility decisions (Kroneberg et al. 2010a, 2010b; Nauck 2007).

While MFS has found considerable support in empirical applications, it has to be kept in mind that it is but one version of DPP. Other advocates of DPP propose models that differ in important respects, in particular regarding the conception of the interplay between Type 1 and Type 2 processes (e.g. Fazio 1990; Wikström 2014). Adjudicating between different models would be eased by adopting experimental designs from cognitive psychology such as overloading working memory, putting subjects under time pressure, as well as priming techniques. Another important avenue is to apply these and related ideas to explain social behavior in real-world settings. For example, the study of Bruch et al. (2016) demonstrates the power of

[11] This does not imply that selections in the as-mode are always and entirely autonomous, fast, based on association or unconscious. In contrast to such a psychological characterization of Type 1 processes, the model asks *in what respects* a behavior is governed by spontaneously activated mental models or based on deliberate choice. For example, some citizens might not even consider to abstain from a national election based on a strongly internalized and activated civic duty norm (as-mode), while others might weigh the perceived costs and benefits of doing so (rc-mode) (Kroneberg 2014). Still, even the former actors certainly do not behave unconsciously and do engage in deliberate behavior in other respects, e.g. which party or candidate to vote for.

Clemens Kroneberg and Andreas Tutic - 9781789906851

new detailed behavioral data sources. Analyzing data on 1.1 million browsing and writing decisions observed on an online dating site, the authors examine mate seeker's decisions about whom to "browse" and, subsequently, whom to write. Both stages involve definitions of the situation (e.g. who is attractive or worthy of attention?) and also exclusion of alternatives based on the (probably spontaneously) application of noncompensatory criteria (e.g. "no one over 40"). It is only thereafter that decision-makers weigh a wider range of attributes and make more deliberate choices.

From a decision-theoretic point of view, the MFS and related Dual-Process theories are purely descriptive decision theories, which lack an axiomatic foundation. This weakness is shared with sociological SEU theory and stems from the different contexts of application of these action theories (e.g. survey research or field experiments). However, there are also attempts to bridge these two lines of action-theoretic traditions. Such a research program applies the axiomatic logic of analysis as well as specific modelling techniques from Bounded Rationality to DPP (Tutić 2015a). Tutić (2015b) has made a first step in this direction by formalizing one critical aspect of MFS in the PCM framework:

Definition 11. A decision-maker exhibiting Script Obedience is characterized by a pair $(\succ, \mathbb{S})$, *in which* $\succ$ *is a quasi-complete strict preference on* X *and* $\mathbb{S} \subseteq X \times \boldsymbol{X}$ *is a system of scripts satisfying the following properties:*

- (*Feasibility*) $(s, \mathcal{S}) \in \mathbb{S} \Rightarrow s \in \mathcal{S}$
- (*Consistency*) $(s, \mathcal{S}) \in \mathbb{S} \Rightarrow (s, T) \in N$ for all $s \in \mathrm{T} \subseteq S$
- (*Non-Contradiction*) $(s, \mathcal{S}) \in \mathbb{S} \Rightarrow (t, \mathcal{S}) \notin \mathbb{S}$ for all $t \in X$, $t \neq s$

Script obedience induces a choice function by letting

$$C^{\succ,\mathbb{S}}(\mathcal{S}) := \begin{cases} s, & (s, \mathcal{S}) \in \mathbb{S}, \\ C^{\succ}(\mathcal{S}), & \text{otherwise.} \end{cases}$$

Here, the system of scripts is used to model the routine grounds of action, which apply to certain choice problems. The decision procedure under consideration is simple: If a prescribed script applies to a choice problem, the decision-maker enacts this script (Type 1 process). If no script applies, the decision-maker maximizes his/her preference relation (Type 2 process).

Tutić (2015b) provides a behavioral characterization of Script Obedience, or an exact description of which kind of behavioral patterns are explainable by this decision procedure. It turns out that Script Obedience can explain behavior patterns which satisfy the following property:

Definition 12. A choice function C *satisfies One Deviation (**OD**) if for all* $S, T \in \boldsymbol{X}$ *with* $C(S) \in T \subseteq S$ *and* $C(T) \neq C(S)$; $C(R) = C(T)$ *for all* $R \subseteq T$ *satisfying* $C(T) \in R$.

Note that the premise of **OD** amounts to violation of **WARP** (see Definition 4). So **OD** demands that once a deviation from orthodox rationality occurs, "within" (in a set-theoretic sense) this deviation no other violation takes place.

Clemens Kroneberg and Andreas Tutic - 9781789906851

As should be expected by now, the proof of this characterization involves a measurement theory for preference relations as well as for system of scripts. In addition, the proof relates Script Obedience to other decision procedures in the literature on precise axiomatic terms. For instance, it is shown that Script Obedience is a special case of PCM with Limited Attention and the exact nature of their differing is pinpointed in behavioral as well as procedural terms (Tutić 2015b: Subsection 3.3).

5. CONCLUSION

Analytical sociology is home to several styles of research and explanatory ambitions that make different uses of action theory. The dominant understanding of analytical sociology aims at "a toolbox of social mechanisms based on different (not necessarily compatible or related) behavioural assumptions, each of them being more or less useful depending on the explanatory task at hand" (Kalter and Kroneberg 2014: 97). Hedström's advocacy of a simple desires-beliefs-opportunities framework (so-called "DBO theory") is symptomatic of this strategic disinterest in more elaborated theories of action. Indeed, modelling the interactions of heterogeneous actors that are embedded in spatial and network structures often requires scholars to rely on simple behavioral rules. However, there is also a strong program of action theory in analytical sociology that seeks differentiated explanations of human behavior and values the integration of knowledge that comes from approaching various substantive fields with a unified action-theoretic core (Esser 2009; Lindenberg 2013; Wikström 2014).

In the search for realistic microfoundations, proponents of a strong program of action theory join other disciplines such as cognitive psychology or behavioral economics. Recent developments suggest a convergence in the recognition of dual processes in cognitive and social psychology, sociology and axiomatic theories of Bounded Rationality. The DPP overcomes two central shortcomings of traditional RCT. First, human cognition and action systematically deviate from the calculi employed by proponents of RTC to model these phenomena. Many of these deviations can be explained theoretically by the influence of Type 1 processes (for an overview, see Stanovich 2011). For example, framing effects, such as the fact that actors behave more pro-socially when a poster with a pair of eyes is in sight (*Watching-Eyes Effect*; cf. Haley and Fessler 2005), can be explained by an automatic definition of the situation. Second, DPP offers a much richer and descriptively more accurate picture of the mental life of actors than traditional RCT. For instance, the MFS (Esser 1996; Kroneberg 2014) attributes to actors not only the mental constructs of preferences and subjective probabilities, but also frames and scripts. Thus, compared to traditional RCT, DPP is clearly more compatible with ideas and concepts of alternative sociological paradigms such as symbolic interactionism, phenomenological sociology and also practice theories (Bourdieu 1990; Giddens 1984). This is already evident in recent developments in American cultural sociology (e.g. Vaisey 2009; Miles 2015), in which DPP is taken as the basis for an integration of two traditionally competing conceptualizations of cultural phenomena—i.e. culture-as-a-toolkit (Swidler 1986) or culture-as-a-motive (Weber 1922/1972; Parsons 1937). Ultimately, the advancement and elaboration of DPP might allow analytical sociology to close the idealistic blank space of the traditional and, as Parsons would say, positivistic-utilitarian RCT, which has always been the central point of critique by proponents of alternative paradigms.

Our brief exposition of analytical approaches identified distinct action-theoretic lines of research that differ in the roles assigned to action theories, the level of formal rigor, as well as the standards of evidence used to evaluate action theories. We think that a key task for future research is to link these threads: For example, one could try to use models of multi-stage decisions, framing and dual-processes in established agent-based models, one could apply the axiomatic method from decision theory to the DPP, or one could test sociological action theories also in lab experiments. Such attempts are likely to generate new insights in the short run and could contribute to a more unified and coherent body of knowledge in the long run.

REFERENCES

Ajzen, I. (1991), "The Theory of Planned Behavior", *Organizational Behavior and Human Decision Processes* **50**: 179–211.

Axelrod, R. (1997), "The Dissemination of Culture: A Model with Local Convergence and Global Polarization", *Journal of Conflict Resolution* **41**: 203–26.

Bolton, G.E. and A. Ockenfels (2000), "ERC: A Theory of Equity, Reciprocity, and Competition, *American Economic Review* **90**: 166–93.

Boudon, R. (2003), "Beyond Rational Choice", *Annual Review of Sociology* **29**: 1–21.

Bourdieu, P. (1990), *The Logic of Practice*, Stanford: Stanford University Press.

Breen, R. and J. Goldthorpe (1997), "Explaining Educational Differentials: Towards a Formal Rational Action Theory", *Rationality and Society* **9**: 275–305.

Bruch, E. and J. Atwell (2015), "Agent-Based Models in Empirical Social Research", *Sociological Methods and Research* **44**: 186–221.

Bruch, E., F. Feinberg and K.Y. Lee (2016), "Extracting Multistage Screening Rules from Online Dating Activity Data", *Proceedings of the National Academy of Sciences of the United States of America* **113**: 10530–5.

Bruch, E. and R.D. Mare (2006), "Neighborhood Choice and Neighborhood Change", *American Journal of Sociology* **112**: 667–709.

Coleman, J.S. (1990), *Foundations of Social Theory*, Cambridge, MA: The Belknap Press.

Elster, J. (1989), "Social Norms and Economic Theory", *Journal of Economic Perspectives* **3**: 99–117.

Elster, J. (2009), "Emotions", pp. 51–71, in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press.

Emirbayer, M. and A. Mische (1998), "What Is Agency?", *American Journal of Sociology* **103**: 962–1023.

Esser, H. (1991), *Alltagshandeln und Verstehen*, Tübingen: Mohr.

Esser, H. (1996), "Die Definition der Situation", *Kölner Zeitschrift für Soziologie und Sozialpsychologie* **48**: 1–34.

Esser, H. (2004), "Does the 'New' Immigration Require a 'New' Theory of Intergenerational Integration?", *International Migration Review* **38**: 1126–59.

Esser, H. (2009), "Rationality and Commitment: The Model of Frame Selection and the Explanation of Normative Action", pp. 207–30, in M. Cherkaoui and P. Hamilton (eds), *Raymond Boudon: A Life in Sociology*, Oxford: The Bardwell Press.

Esser, H. and C. Kroneberg (2015), "An Integrative Theory of Action: The Model of Frame Selection", pp. 63–85 in E.J. Lawler, S.R. Thye and J. Yoon (eds), *Order on the Edge of Chaos: Social Psychology and the Problem of Social Order*, Cambridge: Cambridge University Press.

Evans, J. (2010), *Thinking Twice: Two Minds in One Brain*. Oxford: Oxford University Press.

Evans, J. and K.E. Stanovich (2013), "Dual-Process Theories of Higher Cognition: Advancing the Debate", *Psychological Science* **8**: 223–41.

Fazio, R.H. (1990), "Multiple Processes by which Attitudes Guide Behavior: The MODE Model as an Integrative Framework", *Advances in Experimental Social Psychology* **23**: 75–109.

Fehr, E. and K.M. Schmidt (1999), "A Theory of Fairness, Competition, and Cooperation", *Quarterly Journal of Economics* **114**: 817–68.

Finkel, S.E., E.N. Muller and K.-D. Opp (1989), "Personal Influence, Collective Rationality, and Mass Political Action", *American Political Science Review* **83**: 885–903.

Giddens, A. (1984), *The Constitution of Society*, Cambridge: Polity Press.

Gigerenzer, G. and P.M. Todd (1999), *Simple Heuristics that Make Us Smart*, Oxford: Oxford University Press.

Gintis, H. (2017), *Individuality and Entanglement: The Moral and Material Bases of Social Life*, Princeton, NJ: Princeton University Press.

Goldthorpe, J. (1998), "Rational Action Theory for Sociology", *British Journal of Sociology* **49**: 167–92.

Clemens Kroneberg and Andreas Tutic - 9781789906851

Haley, K.J. and D.M.T. Fessler (2005), "Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game", *Evolution and Human Behavior* **26**: 245–56.
Healy, K. (2017), "Fuck Nuance", *Sociological Theory* **35**: 118–27.
Hedström, P. and P. Bearman (2009), "What is Analytical Sociology All About? An Introductory Essay", pp. 3–24, in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press.
Hedström, P. and R. Swedberg (1998), "Social Mechanisms: An Introductory Essay", pp. 1–31, in P. Hedström and R. Swedberg (eds), *Social Mechanisms*: *An Analytical Approach to Social Theory*, Cambridge: Cambridge University Press.
Hedström, P. and P. Ylikoski (2014), "Analytical Sociology and Rational-Choice Theory", pp. 57–71, in G. Manzo (ed.), *Analytical Sociology: Actions and Networks*, Chichester: Wiley.
Joas, H. and W. Knöbl (2009), *Social Theory*: *Twenty Introductory Lectures*, Cambridge: Cambridge University Press.
Kahneman, D. (2011), *Thinking, Fast and Slow*, London: Penguin.
Kahneman, D. and A. Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk", *Econometrica* **47**: 263–92.
Kalter, F. and C. Kroneberg (2014), "Between Mechanism Talk and Mechanism Cult: New Emphases in Explanatory Sociology and Empirical Research", pp. 91–115, in J. Friedrichs and A. Nonnenmacher (eds), *Soziale Kontexte und soziale Mechanismen. Sonderheft 54 der Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Wiesbaden: Springer VS.
Kenrick, D.T. and V. Griskevicius (2013), *The Rational Animal. How Evolution Made Us Smarter Than We Think*, New York: Basic Books.
Kreps, D.M. (1988), *Notes on the Theory of Choice*, Boulder, CO: Westview Press.
Kroneberg, C. (2005), "Die Definition der Situation und die variable Rationalität der Akteure. Ein allgemeines Modell des Handelns", *Zeitschrift für Soziologie* **34**: 344–63.
Kroneberg, C. (2014), "Frames, Scripts, and Variable Rationality: An Integrative Theory of Action", pp. 97–123, in G. Manzo (ed.), *Analytical Sociology: Actions and Networks*, Chichester: Wiley.
Kroneberg, C. and F. Kalter (2012), "Rational Choice Theory and Empirical Research. Methodological and Theoretical Contributions in Europe", *Annual Review of Sociology* **38**: 73–92.
Kroneberg, C., M. Yaish and V. Stocké (2010a), "Norms and Rationality in Electoral Participation and in the Rescue of Jews in WWII: An Application of the Model of Frame Selection", *Rationality and Society* **22**: 3–36.
Kroneberg, C., I. Heintze and G. Mehlkop (2010b), "The Interplay of Moral Norms and Instrumental Incentives in Crime Causation", *Criminology* **48**: 259–94.
Kruglanski, A.W. and G. Gigerenzer (2011), "Intuitive and Deliberate Judgements Are Based on Common Principles", *Psychological Review* **118**: 97–109.
Lindenberg, S. (1992), "The Method of Decreasing Abstraction", pp. 3–20, in J.S. Coleman and T.J. Fararo (eds), *Rational Choice Theory*: *Advocacy and Critique*, Newbury Park, CA: SAGE.
Lindenberg, S. (2013), "Social Rationality, Self-Regulation, and Well-Being: The Regulatory Significance of Needs, Goals, and the Self", pp. 72–112, in R. Wittek, T.A.B Snijders and V. Nee (eds), *Handbook of Rational Choice Social Research*, Stanford: Stanford University Press.
Lleras, J.S., Y. Masatlioglu, D. Nakajima and E.Y. Ozbay (2017), "When More Is Less: Limited Consideration", *Journal of Economic Theory* **170**: 70–85.
Manzo, G. (2013), "Is Rational Choice Theory *Still* a Rational Choice of Theory? A Response to Opp", *Social Science Information* **52**: 361–82.
Masatlioglu, Y., D. Nakajima and E.Y. Ozbay (2012), "Revealed Attention", *American Economic Review* **102**: 2183–205.
Masatlioglu, Y. and E.A. Ok (2005), "Rational Choice with Status-Quo Bias", *Journal of Economic Theory* **121**: 1–29.
Miles, A. (2015), "The (Re-)Genesis of Values: Examining the Importance of Values for Action", *American Sociological Review* **80**: 680–704.
Nauck, B. (2007), "Value of Children and the Framing of Fertility: Results from a Cross-Cultural Comparative Survey in 10 Societies", *European Sociological Review* **23**: 615–29.
Ok, E.A. (2007), *Real Analysis with Economic Applications*, Princeton, NJ: Princeton University Press.
Opp, K.-D. (1999), "Contending Conceptions of the Theory of Rational Action", *Journal of Theoretical Politics* **11**: 171–202.
Opp, K.-D. (2013), "What Is Analytical Sociology? Strengths and Weaknesses of a New Sociological Research Program", *Social Science Information* **52**: 329–60.
Parsons, T. (1937), *The Structure of Social Action. A Study in Social Theory with Special Reference to a Group of Recent European Writers*, New York: The Free Press.
Read, D. (2007), "Experienced Utility: Utility Theory from Jeremy Bentham to Daniel Kahneman", *Thinking and Reasoning* **13**: 45–61.
Rubinstein, A. (1998), *Modeling Bounded Rationality*, Cambridge, MA: MIT Press.

Rubinstein, A. (2013), "Response Time and Decision Making: A 'Free' Experimental Study", *Judgement and Decision Making* **8**: 540–51.
Rubinstein, A. (2017), *Lecture Notes in Microeconomic Theory*, Princeton, NJ: Princeton University Press.
Rubinstein, A. and L. Zhou (1999), "Choice Problems with a Reference Point", *Mathematical Social Sciences* **37**: 205–9.
Salant, Y. and A. Rubinstein (2008), "(A,f) Choice with Frames", *Review of Economic Studies* **75**: 1287–96.
Samuelson, P.A. (1938), "A Note on the Pure Theory of Consumers' Behavior", *Econometrica* **5**: 61–71.
Sandroni, A. (2011), "Akrasia, Instincts, and Revealed Preferences", *Synthese* **181**: 1–17.
Savage, L.J. (1954), *The Foundations of Statistics*, New York: Dover.
Schelling, T. (1971), "Dynamic Models of Segregation", *Journal of Mathematical Sociology* **1**: 143–86.
Sippel, R. (1997), "An Experiment on the Pure Theory of Consumer's Behaviour", *Economic Journal* **107**: 1431–44.
Stanovich, K.E. (2011), *Rationality and the Reflective Mind*, New York: Oxford University Press.
Swidler, A. (1986), "Culture in Action: Symbols and Strategies", *American Sociological Review* **51**: 273–86.
Tutić, A. (2015a), "Warum denn eigentlich nicht? Zur Axiomatisierung soziologischer Handlungstheorie", *Zeitschrift für Soziologie* **44**: 83–95.
Tutić, A. (2015b), "Revealed Norm Obedience", *Social Choice and Welfare* **44**: 301–18.
Vaisey, S. (2009), "Motivation and Justification: A Dual-Process Model of Culture in Action", *American Journal of Sociology* **114**: 1675–715.
Voss, T. (1985), *Rationale Akteure und soziale Institutionen. Beitrag zu einer endogenen Theorie des sozialen Tauschs*, München: Oldenbourg.
Wakker, P. (2010), *Prospect Theory*: *For Risk and Ambiguity*, Cambridge: Cambridge University Press.
Watts, D. (2014), "Common Sense and Sociological Explanations", *American Journal of Sociology* **120**: 313–51.
Weber, M. [1922] (1972), *Wirtschaft und Gesellschaft*, Tübingen: Mohr.
Wikström, P.-O.H. (2014), "Why Crime Happens: A Situational Action Theory", pp. 74–94, in G. Manzo (ed.), *Analytical Sociology: Actions and Networks*, Chichester: Wiley.

Clemens Kroneberg and Andreas Tutic - 9781789906851

10. Interactions

Michael Mäs

1. INTRODUCTION

Social interaction is the icing on the cake in sociology. In fact, some of the most prominent sociological thinkers argued that social interaction is the core ingredient of dynamics and phenomena that sociologists study. Emile Durkheim proposed, for instance, that sociology's central contribution to the understanding of human behavior is that social interaction can generate collective phenomena that cannot be explained by characteristics of the individuals constituting the collective alone (Cherkaoui 2008; Durkheim 1982). In a nutshell, he argued that "the whole" differs from the "sum of the parts", because the parts interact, which can spark dynamics resulting in unexpected and intriguing sociological phenomena. Also focusing on this puzzling mismatch between individual behavior and collective phenomena, analytical sociologists insist that collective phenomena can and should be explained by theories of interacting individual behavior (Hedström 2005; Hedström and Swedberg 1998; Hedström and Ylikoski 2010). In this chapter, I reflect on alternative ways of representing social interactions in theoretical models and formulate recommendations that guide individual researchers and the whole community of analytical sociologists in the process of understanding the collective effects of interaction.

I use a very broad definition of the term "social interaction": There is an instance of social interaction when an individual responds to an event in his/her social environment by adopting an attribute s/he would not have adopted in the absence of the event. Individual attributes can be beliefs, opinions, or the behavior of an individual. Furthermore, interaction can also affect ascribed attributes of an individual, such as his/her social status. With "social environment", I refer to other individuals that in one way or the other exert influence on the focal actor. These individuals can be persons in the direct network of the focal actor such as friends and family members. However, there can also be interaction between individuals who are unaware of each other. This definition is general in that it is not restricted to specific characteristics of individuals (such as only their behavior) but includes any adjustment of an individual's characteristics. What is more, this definition does not focus on a specific reason an individual adjusted a characteristic, including for instance behavior changes resulting from social pressure as well as behavior changes that are the mere consequence of imitation. Interaction also does not require a conscious decision process. For instance, an individual may adjust his/her opinion about a politician after having debated with a friend. Having learned new arguments about the politician, his/her opinion might shift without actual conscious reflection. Likewise, the individual causing the interaction may be unaware of his/her effect on others. There are, for instance, theories of so-called "descriptive norms," which assume that individuals copy the behavior that they observe frequently (Cialdini, Reno and Kallgren 1990; Mäs and Opp 2016; Schultz et al. 2007). For instance, if a sufficient number of households in a neighborhood installed solar panels on the roofs of their houses, then other neighbors might be inspired to also invest in solar power. The early adopters may not even be aware of the fact their behavior affected the

Michael Mäs - 9781789906851

decisions of others. In the definition of social interaction, I explicitly include cases where an actor decided to not change a belief, opinion, or behavior as a response to another person. For instance, according to the definition it is an act of interaction when a person expecting negative reactions from his/her friends decides to not invest in a new car.

In this chapter, I formulate a set of general recommendations for researchers studying macro phenomena arising from social interaction. To this end, I first reflect on different social processes that lead to social interaction and argue that researchers interested in the macro consequences of social interaction require an accurate model of these processes, as they determine when there is social interaction and how interaction changes actor attributes. Second, I echo the proposal by Lopez-Pintado and Watts that researchers should formulate so-called "influence-response functions", mathematical equations that specify how an individual's attributes are affected by interaction (Lopez-Pintado and Watts 2008). These functions are abstract, as they do not mimic the actual process leading to social interaction but they describe the outcome of interaction: A change in individual characteristics. There are, for instance, agent-based models of social-influence dynamics in networks that make explicit assumptions about how individuals select arguments they share with interaction partners and how their interaction partners contemplate their opinions after having been exposed to a specific argument (Mäs and Flache 2013; Mäs et al. 2007; Urbig and Malitz 2005). Influence-response functions of opinion dynamics resulting from argument communication, in contrast, do not explicitly represent arguments but specify in a single function how the opinion of an agent changes after interaction with another agent (Dandekar et al. 2013; Feliciani et al. 2020; Mäs and Bischofberger 2015).

While influence-response functions are very abstract, they come with two important advantages. First, often even fundamentally different processes underlying interaction translate into very similar influence-response functions. As a consequence, findings about the macro consequences derived from one model can be extended to other models, even when the models are concerned with fundamentally different processes and contexts. Second, when influence-response functions of alternative processes differ, they allow a straightforward comparison between models and help identify those assumptions that have critical impact on macro outcomes. Knowing these critical assumptions guides empirical research, I argue. What is more, influence-response functions can serve as a tool for analytical sociologists to communicate about interaction and its macro effects, bridging long-standing divides between subcommunities favoring alternative theoretical paradigms (Castelfranchi et al. 2006; Hedström and Ylikoski 2014; Lopez-Pintado and Watts 2008; Manzo 2013; Opp 2013; Ylikoski 2013).

Since social interaction is a central concept in sociology, there is much literature on its causes and consequences. There is also a rich literature describing how humans perceive interaction and behave in the process of interaction. I am not covering this vast literature here. In contrast, I limit myself to different literatures that are focused on explaining collective phenomena based on assumptions about individual behavior and that assume, in one way or the other, social interaction as defined. In particular, I build on work from the fields of evolutionary game theory (Foster and Young 1990; Young 2001), models of opinion dynamics (Flache et al. 2017; Friedkin and Johnsen 2011; Mason et al. 2007), complexity research (see Chapter 4 in this book), the agent-based social-simulation literature (Bianchi and Squazzoni 2015; Bonabeau 2002; Gilbert and Troitzsch 1999; Squazzoni 2012), and modeling work by analytical sociologists (Baldassarri and Bearman 2007; Goldberg and

Michael Mäs - 9781789906851

Stein 2018; Manzo and Baldassarri 2015). In these fields, theories of social interaction have been developed to explain macro outcomes as diverse as residential segregation, the emergence of social institutions, the diffusion of political action, opinion polarization, group decision-making, the emergence of status hierarchies, and group discrimination. On the one hand, there seems to be important disagreement between these literatures, in particular when it comes to the micro theories describing social interaction. For instance, game theorists often assume that actors make rationally consistent decisions, a paradigm that is criticized by some authors from the other literatures (Castelfranchi et al. 2006; Hedström and Ylikoski 2014; Manzo 2013). On the other hand, I argue that there is an enormous potential for mutual fertilization between the fields, because often even contradictory micro theories of social interaction translate into very similar formal representations of social interaction in models studying the macro consequences of interaction. As a consequence, while the causes of interaction can be very different, the collective phenomena arising from interaction can be very similar. This suggests that there might be very general models of social interaction that can be used to represent seemingly very different social processes and explain diverse collective phenomena.

2. PROCESSES CAUSING INTERACTION

Interaction can result from many different processes. That is, there are various reasons an individual's attributes change as a response to his/her social environment. A central proposition of this chapter is that theories explaining collective phenomena arising from social interaction need to be informed by what processes have caused the interaction between individuals, as different forms of social interaction can generate fundamentally different macro patterns.

It is important to distinguish causes of interaction because the adjustments in beliefs, opinions, behavior and ascribed attributes can depend on the process motivating them. Often social interaction leads to assimilation in that the interaction partners grow more similar. Sometimes, however, interaction can also lead to differentiation in that interaction leads to intensified differences. The process causing interaction determines whether interaction is assimilative or differentiating and when it is assimilative, differentiating or absent. Reflecting about the micro processes causing interaction helps researchers choose an accurate representation of interaction.

One of the most important causes of social interaction are so-called "externalities", a phenomenon studied in the field of game theory (Coleman 1990; Foster and Young 1990). Individuals cause externalities when their behavior generates consequences for other individuals. Unlike other forms of interaction where individuals seek to change the characteristics of others, externalities are often unintended. For instance, when one of your colleagues decides to switch to another word-processing software, you may consider switching too, as it is easier to collaborate with a colleague using the same tools. You may switch software without any intervention from your colleague. It is enough to know that she is using a specific tool. Likewise, Schelling assumed in his seminal model of residential segregation that individuals seek to live in neighborhoods where their in-group is sufficiently represented. As a consequence, when a household moves to another house, the composition of the old and new neighborhood changes, which creates an externality on the household's old and new neighbors.

Michael Mäs - 9781789906851

Often closely related to externalities is social pressure, another process underlying interaction. When an individual A creates an externality on B, then B has an incentive to exert social pressure on A in order to convince A to behave according to B's preferences (Coleman 1990; Opp 2001; Voss 2001). Empirical research has demonstrated that in social dilemma situations individuals are willing to bear costs in order to punish defectors (Fehr and Gächter 2000). There is also a rich empirical literature on social pressures to conform with group standards (Allport 1954; Asch 1955; Homans 1950; Sherif 1935).

Social-psychological theories emphasize that interaction can result from persuasion processes (Myers 1982; Vinokur and Burnstein 1978), where individuals update opinions, beliefs, or behavior after communication with interaction partners. Interaction partners can provide each other with information and arguments that motivate change. Likewise, theories of cognitive dissonance and cognitive balancing imply that individuals adjust attributes in order to be similar to interaction partners they like. Disagreement with positively evaluated interaction partners leads to cognitive dissonance that individuals seek to resolve according to these theories (Festinger 1957; Heider 1967; Opp 1984).

Various theories propose that imitation and social learning are important processes underlying social interaction. In situations where individuals feel uncertain about their beliefs, opinions and behavior, they may copy characteristics they observe in others. Individuals may imitate characteristics of successful others, other individuals who are sufficiently similar, or adopt the most frequent behavior in their social environment (Durrett and Levin 2005; Mäs and Opp 2016; Schultz et al. 2007; Selten and Apesteguia 2005; Thaler and Sunstein 2009). Social imitation appears to be rooted in the human nervous system (Rizzolatti and Craighero 2004) and has been argued to have evolved as a successful decision-making strategy from an evolutionary process (Richerson and Boyd 2005).

There are, however, also forms of social interaction that lead to deviations from others' characteristics. Theories of social differentiation and inter-group relations, for instance, include the assumption that individuals seek to be distinct from negatively evaluated social categories (Bourdieu 1984; Elias 1969; Tajfel and Turner 1979). Most prominently, social identity theory is based on the assumption that individuals adopt characteristics that deviate from typical characteristics of salient outgroups (Salzarulo 2006; Tajfel and Turner 1986). Sociological theories, in contrast, assume that members of privileged classes in a society develop increasingly elaborated and costly behavioral patterns in order to be different from lower societal classes (Bourdieu 1984; Bryson 1996; Elias 1969). Social-psychological research on "striving for uniqueness" finds that individuals adjust their opinions when they realize that many peers hold similar views (Mäs et al. 2010; Smaldino and Epstein 2015; Vignoles et al. 2000).

In sum, there are various cognitive and social processes that can lead to interaction. Many of them imply that individuals grow more similar as a result of social interaction. Nevertheless, depending on the exact underlying process, individuals may show this positive form of social influence only under certain conditions and may sometimes even intensify differences to interaction partners. These differences between the underlying processes can translate into very different macro dynamics, as I show in the next section. It is, therefore, important that researchers interested in the macro-effects of social interact build their analyses on an accurate description of the process underlying social interaction in the context of their study.

Michael Mäs - 9781789906851

3. INFLUENCE-RESPONSE FUNCTIONS AND THEIR MACRO-CONSEQUENCES

The central argument developed in the previous section is that different processes underlying social interaction can result in very different adjustments of individuals' attributes. The core conclusion drawn from this insight is that researchers studying the effects of social interaction on macro phenomena should carefully explicate and empirically test their assumptions about the process underlying social interaction in the context of their study.

Somewhat counter to the argument that underlying processes need to be carefully specified, I argue in this section that for the study of macro consequences of social interaction, it is desirable to work with an abstract representation of the social interaction rather than models that explicitly represent the process underlying social interaction. That is, modelers should specify and analyze what Lopez-Pintado and Watts denoted as "influence-response functions", mathematical equations that describe the change in an actor's attribute(s) resulting from interaction as a function of the actor's attribute before the interaction and the attribute(s) of the interaction partner(s) (Lopez-Pintado and Watts 2008). That is, these functions do not specify the cognitive and social processes leading to social interaction, but they describe the outcome of an interaction, the change in interaction partners' attributes.

One of the most prominent influence-response functions is certainly Granovetter's (1978) seminal threshold model. Granovetter was interested in explaining the dynamics of collective behavior like political protests or riots and argued that the process underlying social interaction results from rational decision-making by individuals. Individuals who consider joining a riot, for instance, will rationally weigh the costs and benefits of participating. The more actors have joined a riot already, the smaller are the costs of participation, as the chances of being apprehended decrease. However, individuals may differ in the degree to which they require safety before entering a riot. Furthermore, the benefits individuals derive from rioting differ, leading Granovetter to the conclusion that every individual can be described by a threshold that quantifies the number or share of other individuals who must have joined the collective behavior before the individual also decided to enter. In the mathematical model, Granovetter refrained from implementing individuals' safety preferences, perceived benefits, and the actual weighing of cost and benefits. Using a simple influence-response function, he only included a threshold value for every individual and assumed that individuals would enter the riot when the number of active individuals exceeds this threshold. Thus, an individual changes his/her behavior from inaction to entering a riot when the number of participants rises above his/her individual threshold value.

It turns out that the decision-making process in a very different social context leads to the same influence-response function as assumed by Granovetter (Easley and Kleinberg 2010: 19.2; Lopez-Pintado and Watts 2008). This second example comes from evolutionary game theory and focusses on coordination games in networks. Assume, as an example, that a population of researchers chooses an operating system for their computers. There are two options, PCs and Macs. Furthermore, every researcher has at least one collaborator they seek to coordinate with on a joint system, as working with different systems is highly inefficient. In a nutshell, every pair of collaborators is playing the coordination game shown in Table 10.1. If two collaborators work with a PC, they both receive a payoff of a. Both get b, if both use a Mac. The values of a and b might be the same for every actor but they might also differ when some researchers happen to enjoy working with one of the options more than with the other.

Michael Mäs - 9781789906851

*Table 10.1 Coordination game (*a *and* b *exceed 0)*

		Researcher 1	
		PC	Mac
Researcher 2	PC	*a*,*a*	0,0
	Mac	0,0	*b*,*b*

When two collaborators work with different systems, however, they receive a payoff of zero, which implements that this is an undesirable outcome.

Suppose a given scientist has d collaborators and a fraction, p, of them is using a PC and the remainder use Macs $(1 - p)$. What system should the researcher choose? If the researcher chooses a PC, his/her payoff will be $p \times d \times a$. Buying a Mac will lead to a payoff of $(1 - p) \times d \times b$. Thus, a PC is the better choice whenever $p \times d \times a > (1 - p) \times d \times b$. Rearranging terms leads to the conclusion that the researcher will choose a PC when

$$p > \frac{b}{a+b}. \tag{10.1}$$

Equation 10.1 specifies for every researcher a threshold depending on his/her personal a and b values. When the share of collaborator exceeds the term on the right-hand-side of Equation 10.1, then the researcher will choose a PC. Thus, the influence-response function in this second example turns out to be the same threshold model as assumed by Granovetter, although it is derived in a very different context and a very different decision problem. This implies that when one studies coordination problems in networks, it is actually not necessary to model the game-theoretical decision-making process of each individual researcher. It suffices to calculate every researcher's threshold and compare it to their current value of p.

A third example shows that the threshold model is even more general. This example comes from the rich literature that emerged from Axelrod's seminal model of the dissemination of culture (Axelrod 1997; Huckfeldt et al. 2004). Axelrod sought to understand the conditions under which a population of agents falls apart into subgroups with different cultural attributes. To this end, he studied a population of agents described by a set of nominal characteristics that are open to social influence. Interaction was modeled as a dyadic event where one agent copies a characteristic of one of his/her neighbors. Extending Axelrod's work, Flache and Macy compared this dyadic form of interaction to the case when influence is social in that individual consider the distribution in cultural traits in their whole ego network when they update cultural traits (Flache and Macy 2011b). One way to implement this is to assume that agents decide to change a trait only when a sufficient share of their interaction partners has adopted that trait. This assumption also leads to the same influence-response function as Granovetter assumed, although it describes a very different process of interaction. Other models using an influence-response function similar to Granovetter's threshold model can be found for instance in the literature on the spreading of norm violation (Mäs and Opp 2016) and complex contagion (Centola and Macy 2007).

These examples illustrate one of the central arguments of the present chapter (Lopez-Pintado and Watts 2008). Although there can be fundamental differences between different

Michael Mäs - 9781789906851

forms of interaction, it is possible that they imply the same or very similar influence-response functions. As a consequence, these processes also generate the same macroscopic patterns despite the fundamental differences on the micro level. This shows that insights about macro effects derived for a specific context can be extended to a seemingly fundamentally different context if the processes underlying social interaction in the two contexts imply the same influence-response function.

What is more, when interaction processes imply different influence-response functions, these abstract representations of interaction turn out to be very useful tools for analysing and understanding the consequences of social interaction. This can be illustrated with models of social influence in networks. These models allow one to study the macro consequences arising from social influence on continuous attributes, such as opinions, and have been used to develop hypotheses about the conditions under which populations fall apart into subgroups with different opinions or even experience opinion polarization (Flache et al. 2017; Friedkin and Johnsen 2011; Mason et al. 2007). In many of these models, every individual is described by a continuous variable that can adopt values between zero and one to represent the stance on a given issue.

One of the most prominent influence-response functions in this realm is the averaging assumption (Flache et al. 2017; French 1956; Friedkin and Johnsen 2011). It assumes that the opinions of interaction partners grow more similar. Equation 10.2 is the resulting influence-response function when an actor *i* interacts with *n* other actors (Flache et al. 2017). The first part of the equation specifies the opinion $o_{i,t+1}$ that agent *i* holds after the interaction (time point *t*+1) is a function of *i*'s initial opinion (time point *t*) and the opinion change $\Delta o_{i,t+1}$ resulting from the social interaction. The second part specifies this opinion change. It is assumed that actors consider the opinions of their *n* interaction partners *j* and adjust their own opinion in the direction of the weighted average of their interaction partners' views. Parameter μ quantifies how open actor *i* is to influence. The weights w_{ij} describe the relative impact that each interaction partners *j* has on *i*. Furthermore, weights are assumed to be positive and to sum up to a value of one. Different weights might represent that some interaction partners are considered more trustworthy or that they are close friends and therefore can exert more influence on *i*. Weights have also been used to implement social power and leadership relations between individuals.

$$o_{i,t+1} = o_{i,t} + \Delta o_{i,t+1} = o_{i,t} + \mu \sum_{j=1}^{n} w_{ij} \left(o_{jt} - o_{it} \right) \tag{10.2}$$

The influence-response function specified in Equation 10.2 has also been labeled the "assimilative social-influence model", as it captures the notion that social interaction leads to assimilation (Flache et al. 2017). This is independent of the micro process causing the interaction. That is, it does not matter whether individuals grow more similar because of persuasion, social pressure, imitation or another process. The equation describes any interaction where interaction partners decrease differences according to the averaging principle. The top graph of Panel a in Figure 10.1 visualizes how opinion shifts depend on the initial differences between two interaction partners. The grey zone shows the range of possible opinion shifts given the initial opinion differences. Positive shifts show that agent *i* moves in the direction of agent *j*. Negative shifts indicate that two agents increase differences as a result of the interaction. The bold black line shows how opinions shift according to Equation 10.2.

The collective consequences arising from social interaction according to the assimilative influence-response function have puzzled researchers for decades. It has been demonstrated

Michael Mäs - 9781789906851

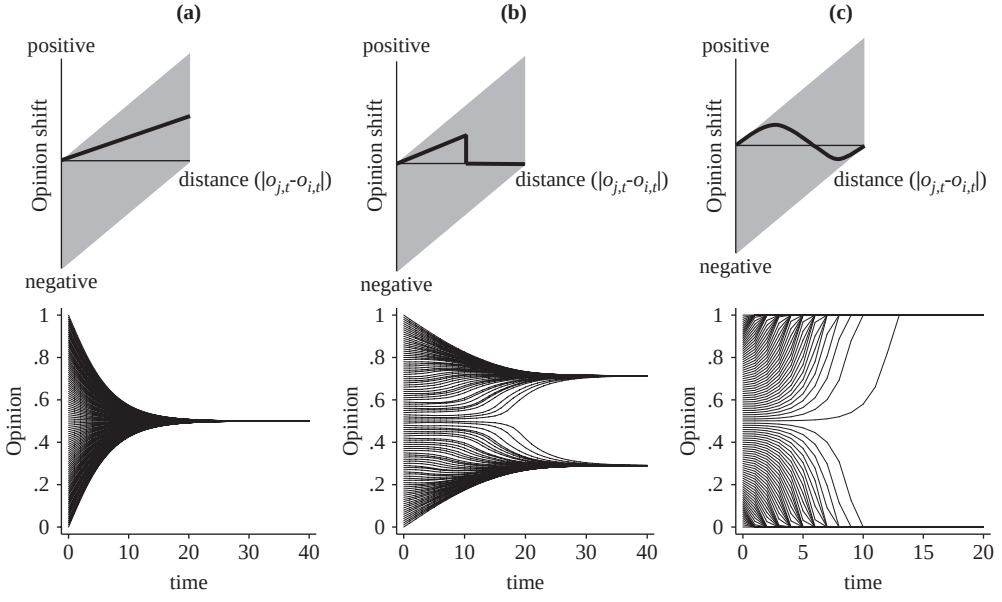


Figure 10.1 Three influence-response functions (top) and the opinion dynamics they generate (bottom)

that a population of individuals interacting according to Equation 10.2 will experience opinion convergence and consensus eventually. Unless the population consists of perfectly unconnected subnetworks with zero interaction between them, consensus is inevitable. On the one hand, this prediction is very plausible, as actors updating their views according to the assimilative influence-response function will always grow more similar to their interaction partners. On the other hand, this observation is puzzling, as it contradicts the dynamics of opinion fragmentation and polarization observed in many empirical studies (Abelson 1964; Axelrod 1997; Flache et al. 2017). Since social influence is a very strong force in social systems, modelers already wondered in the 1960s how models' assumptions need to be adjusted in order to be able to explain the emergence and stability of opinion diversity (Abelson 1964).

An alternative influence-response function that has been developed to resolve this puzzle adds that individuals do not adjust their attributes in the direction of another actor when the differences between the two are too big (Axelrod 1997; Carley 1991; Deffuant et al. 2000; Hegselmann and Krause 2002; Mark 1998). This can result from various underlying processes. Individuals may, for instance, select interaction partners who are sufficiently similar and, therefore, avoid any exposure to dissimilar interaction partners (Feld 1982; McPherson et al. 2001). Alternatively, individuals may consider dissimilar contacts unreliable sources and, therefore, ignore their views when they are exposed to them. The consequence in terms of the attribute change resulting from interaction is the same and can be summarized in the influence-response function specified in Equations 10.3 and 10.4. These equations assume that actor i updates his/her opinion only based on the opinion of a single agent, j. The term $f_w(o_{it},o_{jt})$ in Equation 10.3, which is detailed in Equation 10.4, specifies that there is no

Michael Mäs - 9781789906851

interaction (zero opinion change) when the opinions of agents *i* and *j* differ by more than the threshold value ε.

$$o_{i,t+1} = o_{i,t} + \Delta o_{i,t+1} = o_{i,t} + f_w(o_{jt}, o_{it})(o_{jt} - o_{it}) \tag{10.3}$$

$$f_w(o_{jt}, o_{it}) = \begin{cases} \mu, & if \ |o_{jt} - o_{it}| \leq \varepsilon \\ 0, & otherwise \end{cases} \tag{10.4}$$

Thus, the influence-response function specified in Equations 10.3 and 10.4 keeps the averaging assumption, but adds that the influence weights, which were fixed values in the assimilative influence-response function, adopted a value of μ when actors agree sufficiently or zero otherwise (see Panel b in Figure 10.1). Accordingly, it has been labeled the similarity-biased influence-response function.

The macro-pattern arising from this small change of the influence-response function is very different, as Panel b in Figure 10.1 illustrates. When the value of parameter ε is sufficiently high, even a perfectly connected population where all actors could interact with each other will fall apart into two subgroups will different opinions (Deffuant et al. 2000; Hegselmann and Krause 2002; Kurahashi-Nakamura et al. 2016; Lorenz 2007). In other words, the opinion distribution is fragmenting into multiple peaks. Individuals with opinions between the two peaks will initially hardly adjust their opinions, as they are "pulled" with equal forces into opposite directions. However, at some moment one of the two forces will slightly dominate and these individuals will join one of the subgroups. The emerging pattern of fragmentation is a rest point when agents belonging to the same subgroup perfectly agree and the opinion differences between the subgroups exceed the value of parameter ε.

A third prominent influence-response function is summarized in Equations 10.3 and 10.5. It is also based on the averaging principle but adds that individuals distance themselves from interaction partners who disagree too much. This negative social influence has been derived, for instance, from social-psychological theories of group differentiation (Salzarulo 2006), social judgment theory (Huet and Deffuant 2010; Huet et al. 2008), and theories of cognitive dissonance (Baldassarri and Bearman 2007; Flache and Macy 2011a; Flache and Mäs 2008; Macy et al. 2003; Mark 2003). Independent of the exact underlying process, these models share the assumption that too-big opinion differences result in a repulsive force and increasing opinion differences between interaction partners. Formally, this can be implemented with Equation 10.3 and a different specification of the term $f_w(o_{it},o_{jt})$, as Equation 10.5 shows. Equation 10.5 implements the same weighted averaging as in the other two influence-response functions, but adds that the weights turn negative when opinion differences exceed half of the opinion scale's range. To prevent opinions from crossing the bounds of the opinion scale, many models add a smoothing (Flache and Macy 2011a; Flache and Mäs 2008). That is, the closer the opinion of an agent is to one of the boundaries of the scale, the smaller the moves toward the respective bound.

$$f_w(o_{jt}, o_{it}) = \mu(1 - 2|o_{jt} - o_{it}|) \tag{10.5}$$

Panel c in Figure 10.1 shows how opinion differences translate into opinion shifts according to this influence-response function. Unlike the other two functions, this so-called "repulsive

Michael Mäs - 9781789906851

influence-response function" results in negative opinion shifts. The repulsive influence-response function also generates a very different macro pattern than the other two models (see panel c in Figure 10.1). Unlike the other two functions, it can generate opinion polarization in populations with sufficient interaction between disagreeing agents (Flache and Macy 2011a; Flache and Mäs 2008). When individuals with opinions close to the extremes of the opinion spectrum interact with like-minded others, opinion variance will tend to decrease. However, when there is also interaction between individuals with opposing views, these individuals' opinions will grow more extreme. Subsequently, these agents can pull their like-minded but less extreme interaction partners closer to the poles of the scale, a process that ends when two maximally extreme subgroups have formed.

Comparison of the three panels in Figure 10.1 demonstrates how useful influence-response functions are when it comes to model comparison. All three functions share that interaction between similar individuals leads to decreasing opinion differences between interaction partners. Adding the assumption that there is no interaction when individuals differ too much changes macro predictions compared to the assimilative influence-response function. With this additional ingredient, the averaging assumption is able to generate opinion fragmentation. When it is further included that too-big opinion differences cause repulsive opinion changes, the averaging assumption is even able to explain the emergence of opinion polarization.

While a recent review of the social-influence modeling literature concluded that these three functions are the most prominent models in this literature, there are many more candidates (Flache et al. 2017). For instance, modelers included findings from empirical research suggesting that individuals adjust their opinions when they learn that they share the same views as many others. Wanting to be unique, individuals can also be negatively influenced by similar interaction partners (Imhoff and Erb 2009; Vignoles et al. 2000). This form of influence generates very different collective dynamics than the "repulsive influence-response function" (Mäs et al. 2010; Smaldino and Epstein 2015). Another set of models assume reinforcement of actor attributes when individuals with similarly biased characteristics interact. When, for instance, individuals with the same political orientation meet and communicate the arguments underlying their opinions, they might provide each other with new reasons for an even more extreme standpoint (Mäs and Flache 2013). Opinion reinforcement can also result from processes of biased assimilation and reinforcement learning (Banisch and Olbrich 2019; Dandekar et al. 2013). When acting in tandem with homophily, the tendency to mainly communicate with like-minded individuals, reinforcement can also generate opinion polarization, even when there is no negative influence.

4. SUMMARY AND CONCLUSION

Departing from a very general definition of the concept of social interaction, I first reflected on different processes underlying social interaction and showed that different processes can imply markedly different changes in individual attributes. Next, I echoed a recommendation by López-Pintado and Watts who proposed that researchers interested in the effects of interaction between individuals on macro dynamics should formulate what they called "influence-response functions" and illustrated their usefulness in the process of understanding the macro consequences of assumptions about interaction. Based on these observations, I

Michael Mäs - 9781789906851

formulate in the following a set of recommendations that will help analytical sociologists in their research on the micro-macro problem.

Recommendation 1: Describe in detail the process underlying social interaction in the context of your study.

As I have tried to illustrate, there are various reasons individual characteristics change during interaction. The magnitude and the direction of change can greatly differ across different forms of interaction and different contexts. In the process of aggregation from the micro to the macro level even small inaccuracies in the description of the micro process can generate important deviations between the macro-predictions of theories and real-life collective behavior. It is, therefore, critical that researchers interested in the effects of interaction on macro dynamics build their analyses on a detailed model of social interaction. That is, it needs to be specified how individuals interact. It can matter critically, for instance, whether an individual interacts with a single interaction partner or whether changes in attributes depend on influence by multiple actors (Flache and Macy 2011b). Likewise, it can matter whether an individual exerts influence on a single actors or whether the person exerts the same influence on multiple actors at the same time (Keijzer et al. 2018). Next, it needs to be specified why individual characteristic change during interaction and what between-individual process is responsible (e.g. persuasion, external effects or social pressure).

Recommendation 2: Translate the process underlying social interaction into an influence-response function and compare the result with alternative functions that may be relevant in the context of your study.

To be sure, I do not propose in the second recommendation that models making explicit the process underlying interaction are problematic. In fact, these models can be very useful because they are more general than influence-response functions. Granovetter's threshold model, for instance, follows from the theory of rational action in a context where decision-makers have exactly two options. When there are more action alternatives, the same micro theory implies a different influence-response function. Influence-response functions, in other words, depend on the context and are, thus, so-called "middle range theories" (Boudon 1981; Hedström and Udehn 2009; Hedström and Ylikoski 2010; Merton 1957). Theories describing the process underlying interaction, in contrast, can often be applied to various contexts. Likewise, Recommendation 2 does not conflict with the increasingly frequent calls for cognitively realistic models (Conte and Giardini 2016; Hedström and Ylikoski 2011, 2014; Jager 2017). Also, influence-response functions should accurately describe how individual attributes change during interaction, reflecting all aspects of attribute change that critically affect macro-outcomes. This description, however, is condensed to a single function rather than a set of assumptions describing the process leading to attribute changes. It is for this reason that Recommendation 1 is important.

For the purpose of modeling and understanding the macro consequences of interaction, influence-response functions come with important advantages compared to models explicitly representing the processes underlying social interaction. In particular, Lopez-Pintado and Watts argued that influence-response functions are very powerful tools as they allow one to extend results from one class of models to other classes even when the models are concerned

Michael Mäs - 9781789906851

with fundamentally different phenomena or social contexts (Lopez-Pintado and Watts 2008). If the influence-response function of two models is sufficiently similar, then the models will make identical macro predictions. As a consequence, insights about the macro outcomes derived with one model also apply to the other model.

Even within the relatively homogenous community of analytical sociologists, there are long debates about alternative micro-theories. While there is, for instance, a strong subcommunity of modelers studying rational decision-making and game-theoretical models, many analytical sociologists argue that the theory of rational action is fundamentally flawed and should, therefore, be replaced by models of action that assume decision-making based on heuristics (Castelfranchi et al. 2006; Hedström and Ylikoski 2014; Manzo 2013; Opp 2013; Ylikoski 2013). While these different theoretical micro approaches seem to be fundamentally different, influence-response functions promise to serve as a communication device between the fields and a method to transfer insights from one subcommunity to the other. This leads to the next recommendation:

Recommendation 2a: If it turns out that the influence-response function you arrived at is identical or very similar to functions used in other fields studying social interaction, explore whether findings about the macro consequences derived in other fields are also relevant for the context of your study.

There are various fields where analytical sociologists might find models with similar influence-response functions. A great candidate is certainly the rich literature from evolutionary game theory, a field that is also focused on micro-macro transitions and studies both rational decision-making and simpler decision rules (Foster and Young 1990; Pradelski and Young 2012; Young 2001; Young 2015). But there are many more disciplines that are concerned with micro-macro problems where interaction plays a central role, including fields as diverse as computer science, physics, mathematics, and biology (see Chapter 4 in this book).

Recommendation 2b: If it turns out that your influence-response function differs from other functions, identify the crucial differences between functions. That is, conduct theoretical analyses comparing the macro-predictions of alternative influence-response functions to understand what differences in the influence-response functions cause the variation in macro-predictions.

Obviously, one will very often arrive at the conclusion that an influence-response function is unique. However, also in this case, influence-response functions are powerful tools, as they allow one to compare the macro-predictions with the predictions of similar but not identical functions. If it turns out that the macro predictions do not differ in important ways, one will conclude that the differences are innocent, which is also a very interesting result. If there are, however, important differences between the macro-predictions of one's model and existing theories, then one can draw conclusions about why the models make different macro-predictions. In the above summary of findings from the social-influence literature, for instance, we saw that small differences between the three influence-response functions translate into markedly different macro-dynamics. Repulsive influence, for instance, plays a crucial role in explanations of opinion polarization, since the models that do not assume repulsive influence fail to explain opinion polarization. Thus, comparing a given influence-response function with alternative models of interaction demonstrates what aspects of the function are responsible for

Michael Mäs - 9781789906851

the unique macro predictions of the model. The next step is then to identify those aspects of the underlying process of interaction that imply the critical aspects of the influence-response function. These aspects should then be studied empirically, as they are the critical assumptions of the theory. Identifying these critical assumptions of a theory is much harder without influence-response functions, as theories leading to different functions often differ in many ways, which makes comparing models based on their micro-assumptions very challenging.

These considerations lead to two recommendations for empirical research.

Recommendation 3: Test empirically whether your influence-response function is an accurate description of attribute changes in the context of your study. Focus the test on those aspects of the function that have critical impact on macro-predictions.

While it is of course important to conduct direct tests of the micro-model underlying social interaction, the influence-response function also needs to be tested empirically. In the past years, for instance, advances in so-called "stochastic actor-oriented modelling" have made it possible to estimate influence-response functions from longitudinal network data (Snijders et al. 2010; Steglich et al. 2010). Another approach is to conduct controlled experiments in the laboratory and the field (Clemm von Hohenberg et al. 2017; Takács et al. 2016). The empirical test should be focused on those aspects of the influence-response functions that turned out to be responsible for the differences between the functions' macro-implications and the implications of alternative functions (Takács et al. 2016).

Recommendation 4: Test empirically the macro-predictions of your influence-response function, paying special attention to those predictions that differ from predictions of alternative influence-response functions.

I tried to demonstrate that influence-response functions are great tools for analytical sociologists interested in understanding the macro consequences of individual behavior when individuals do not act in isolation but interact. These abstract representations of interaction have the potential to help analytical sociologists identify findings from other fields that they can adopt in their own research. What is more, they allow one to compare one's model with alternative theories and identify those model assumptions and predictions that require special attention in empirical research. However, influence-response functions are not only a theoretical tool. In addition, they are also a new form of interaction between analytical sociologists, as they allow analytical sociologists to communicate about their models' assumptions and macro predictions even when these models are based on fundamentally different micro-theories, and when they are concerned with seemingly unrelated contexts. Influence-response functions may, thus, be a new unifying paradigm for analytical sociologists.

REFERENCES

Abelson, Robert P. (1964), "Mathematical Models of the Distribution of Attitudes Under Controversy", pp. 142–60, in N. Frederiksen and H. Gulliksen (eds), *Contributions to Mathematical Psychology*, New York: Rinehart Winston.

Allport, Gordon W. (1954), *The Nature of Prejudice*. Boston: Beacon Press.

Asch, S.E. (1955), "Opinions and Social Pressure", *Readings about Soc. Anim.* **193**: 17–26.

Michael Mäs - 9781789906851

Axelrod, R. (1997), "The Dissemination of Culture: A Model with Local Convergence and Global Polarization", *J. Conflict Resolut.* **41** (2): 203–26.
Baldassarri, Delia and Peter Bearman (2007), "Dynamics of Political Polarization", *Am. Sociol. Rev.* **72**: 784–811.
Banisch, S. and E. Olbrich (2019), "Opinion Polarization by Learning from Social Feedback", *J. Math. Sociol.* **43** (2): 76–103.
Bianchi, Federico and Flaminio Squazzoni (2015), "Agent-Based Models in Sociology", *Wiley Interdiscip. Rev. Comput. Stat.* **7** (4): 284–306.
Bonabeau, E. (2002), "Agent-Based Modeling: Methods and Techniques for Simulating Human Systems", *Proc. Natl. Acad. Sci. USA*, **99**: 7280–7.
Boudon, Raymond (1981), *The Logic of Social Action*: *An Introduction to Sociological Analysis*, London: Routledge and Kegan Paul.
Bourdieu, Pierre (1984), *Distinction: A Social Critique of the Judgment of Taste*. Cambridge, MA: Harvard University Press.
Bryson, B. (1996), "'Anything but Heavy Metal': Symbolic Exclusion and Musical Dislikes", *Am. Sociol. Rev.* **61** (5): 884–99.
Carley, Kathleen (1991), "A Theory of Group Stability", *Am. Sociol. Rev.* **56** (3): 331–54.
Castelfranchi, Cristiano, Francesca Giardini, and Francesca Marzo (2006), "Symposium on 'Cognition and Rationality: Part I' Relationships between Rational Decisions, Human Motives, and Emotions", *Mind Soc.* **5** (2):173–97.
Centola, Damon and Michael Macy (2007), "Complex Contagions and the Weakness of Long Ties", *Am. J. Sociol.* **113** (3): 702–34.
Cherkaoui, Mohamed (2008), *Durkheim and the Puzzle of Social Complexity*. Oxford: Bardwell Press.
Cialdini, Robert B., Raymond R. Reno, and Carl A. Kallgren (1990), "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places", *J. Pers. Soc. Psychol.* **58** (6): 1015–26.
Clemm von Hohenberg, Bernhard, Michael Mäs and Bary S.R. Pradelski (2017), *Micro Influence and Macro Dynamics of Opinion Formation*. https://ssrn.com/abstract=2974413.
Coleman, James S. (1990), *Foundations of Social Theory*. Cambridge, MA: Harvard Univ. Press.
Conte, Rosaria and Francesca Giardini (2016), "Towards Computational and Behavioral Social Science", *Eur. Psychol.*, **21** (2).
Dandekar, Pranav, Ashish Goel and David T. Lee (2013), "Biased Assimilation, Homophily, and the Dynamics of Polarization", *Proc. Natl. Acad. Sci. USA*, **110** (15): 5791–96.
Deffuant, Guillaume, David Neau, Frederic Amblard and Gerard Weisbuch (2000), "Mixing Beliefs among Interacting Agents", *Adv. Complex Syst.* **3**: 87–98.
Durkheim, Emile (1982), *The Rules of Sociological Method*, edited by S. Lukes. New York: The Free Press.
Durrett, Richard and Simon A. Levin (2005), "Can Stable Social Groups Be Maintained by Homophilous Imitation Alone?", *J. Econ. Behav. Organ.* **57**: 267–86.
Easley, D. and J. Kleinberg (2010), *Networks, Crowds, and Markets*. New York: Cambridge University Press.
Elias, Norbert (1969), *The Civilizing Process, Vol. I. The History of Manners*. Oxford: Blackwell.
Fehr, Ernst and Simon Gächter (2000), "Cooperation and Punishment in Public Goods Experiments", *Am. Econ. Rev.* **90** (4): 980–94.
Feld, S.L. (1982), "Social Structural Determinants of Similarity among Associates", *Am. Sociol. Rev.* **47** (6): 797–801.
Feliciani, Thomas, Andreas Flache and Michael Mäs (2020), "Persuasion without Polarization? Modelling Persuasive Argument Communication in Teams with Strong Faultlines", *Comput. Math. Organ. Theory* **27** (1): 1–32.
Festinger, L. (1957), *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
Flache, Andreas and Michael Mäs (2008), "How to Get the Timing Right: A Computational Model of the Effects of the Timing of Contacts on Team Cohesion in Demographically Diverse Teams", *Comput. Math. Organ. Theory* **14** (1): 23–51.
Flache, Andreas and M.W. Macy (2011a), "Small Worlds and Cultural Polarization", *J. Math. Sociol.* **35** (1): 146–76.
Flache, Andreas and M.W. Macy (2011b), "Local Convergence and Global Diversity: From Interpersonal to Social Influence", *J. Conflict Resolut.* **55** (6): 970–95.
Flache, Andreas et al. (2017), "Models of Social Influence: Towards the Next Frontiers", *Jasss-the J. Artif. Soc. Soc. Simul.* **20** (4). http://jasss.soc.surrey.ac.uk/20/4/2.html.
Foster, D. and P. Young (1990), "Stochastic Evolutionary Game Dynamics", *Theor. Popul. Biol.* **38** (2): 219–32.
French, J.R.P. (1956), "A Formal Theory of Social Power", *Psychol. Rev.* **63** (3): 181–94.
Friedkin, N.E. and E.C. Johnsen (2011), *Social Influence Network Theory*. New York: Cambridge University Press.
Gilbert, Nigel and Klaus G. Troitzsch (1999), *Simulation for the Social Scientist*. Buckingham: Open University Press.
Goldberg, Amir and Sarah K. Stein (2018), "Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation", *Am. Sociol. Rev.* **83** (5): 897–932.
Granovetter, M. (1978), "Threshold Models of Collective Behavior", *Am. J. Sociol.* **83** (6):1420–43.
Hedström, Peter (2005), *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.

Michael Mäs - 9781789906851

Hedström, Peter and Richard Swedberg (1998), *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
Hedström, P. and L. Udehn (2009), "Analytical Sociology and Theories of the Middle Range", pp. 25–47 in P. Hedström and P. Bearman (eds), *The Oxford Handbook*, Oxford University Press.
Hedström, Peter and Petri Ylikoski (2010), "Causal Mechanisms in the Social Sciences", *Annu. Rev. Sociol.* **36** (1): 49–67.
Hedström, Peter and Petri Ylikoski (2011), "Analytical Sociology", pp. 386–98 in *The SAGE Handbook of the Philosophy of Social Sciences*, London: SAGE.
Hedström, Peter and Petri Ylikoski (2014), "Analytical Sociology and Rational-Choice Theory", pp. 53–70 in G. Manzo (ed.), *Analytical Sociology*, Chichester, United Kingdom: Wiley.
Hegselmann, Rainer and Ulrich Krause (2002), "Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation", *J. Artif. Soc. Soc. Simul.* **5** (3).
Heider, Fritz (1967), "Attitudes and Cognitive Organization", pp. 39–41, in M. Fishbein (ed.), *Readings in Attitude Theory and Measurement*. New York, London, Sydney: Wiley.
Homans, George C. (1950), *The Human Group*. New York: Harcourt, Brace & Co.
Huckfeldt, Robert, Paul E. Johnson and John Sprague (2004), *Political Disagreement*: *The Survival of Diverse Opinions within Communication Networks*. Cambridge: Cambridge Univ. Press.
Huet, Sylvie and Guillaume Deffuant (2010), "Openness Leads to Opinion Stability and Narrowness to Volatility", *Adv. Complex Syst.* **13** (3).
Huet, Sylvie, Guillaume Deffuant and Wander Jager, (2008), "A Rejection Mechanism in 2D Bounded Confidence Provides More Conformity", *Adv. Complex Syst.* **11** (4): 529–49.
Imhoff, R. and H. P. Erb (2009), "What Motivates Nonconformity? Uniqueness Seeking Blocks Majority Influence", *Personal. Soc. Psychol. Bull*. **35** (3): 309–20.
Jager, Wander (2017), "Enhancing the Realism of Simulation (EROS): On Implementing and Developing Psychological Theory in Social Simulation", *J. Artif. Soc. Soc. Simul.* **20** (3): 14.
Keijzer, Marijn A., Michael Mäs and Andreas Flache (2018), "Communication in Online Social Networks Fosters Cultural Isolation", *Complexity* 1–20.
Kurahashi-Nakamura, Takasumi, Michael Mäs and Jan Lorenz (2016), "Robust Clustering in Generalized Bounded Confidence Models", *J. Artif. Soc. Soc. Simul.* **19** (4): 7.
Lopez-Pintado, D. and D.J. Watts (2008), "Social Influence, Binary Decisions and Collective Dynamics", *Ration. Soc.* **20** (4): 399–443.
Lorenz, J. (2007), "Continuous Opinion Dynamics under Bounded Confidence: A Survey", *Int. J. Mod. Phys. C* **18** (12): 1819–38.
Macy, Michael Walton, James A. Kitts, Andreas Flache and Stephen Benard (2003), "Polarization in Dynamic Networks: A Hopfield Model of Emergent Structure", edited by R. Breiger, K. Carley and P. Pattison, *Dyn. Soc. Netw. Model. Anal.*: 162–73.
Manzo, Gianluca (2013), "Is Rational Choice Theory *Still* a Rational Choice of Theory? A Response to Opp", *Soc. Sci. Inf.* **52** (3): 361–82.
Manzo, Gianluca and Delia Baldassarri (2015), "Heuristics, Interactions, and Status Hierarchies", *Sociol. Methods Res.* **44** (2): 329–87.
Mark, Noah P. (1998), "Beyond Individual Differences: Social Differentiation from First Principles", *Am. Sociol. Rev.* **63** (3): 309–30.
Mark, Noah P. (2003), "Culture and Competition: Homophily and Distancing Explanations for Cultural Niches", *Am. Sociol. Rev.* **68** (3): 319–45.
Mäs, Michael and Andreas Flache (2013), "Differentiation without Distancing. Explaining Opinion Bi-Polarization without Assuming Negative Influence", *PLoS One* **8** (11).
Mäs, M., Andreas Flache and D. Helbing (2010), "Individualization as Driving Force of Clustering Phenomena in Humans", *PLoS Comput. Biol.* **6** (10).
Mäs, Michael, Andreas Flache and Károly Takács (2007), "The Double Edge of Demographic Faultlines. The Neglected Effects of Crisscrossing on Cohesion and Performance in Work Teams", Conference presentation.
Mäs, M. and K.D. Opp (2016), "When Is Ignorance Bliss? Disclosing True Information and Cascades of Norm Violation in Networks", *Soc. Networks* **47**.
Mäs, Michael and Lukas Bischofberger (2015), "Will the Personalization of Online Social Networks Foster Opinion Polarization?", *SSRN Electron. J.* Retrieved August 20, 2015, http://papers.ssrn.com/abstract=2553436.
Mason, W.A., F.R. Conrey and E.R. Smith (2007), "Situating Social Influence Processes: Dynamic, Multidirectional Flows of Influence within Social Networks", *Personal. Soc. Psychol. Rev.* **11** (3): 279–300.
McPherson, Miller, Lynn Smith-Lovin and James M. Cook (2001), "Birds of a Feather: Homophily in Social Networks", *Annu. Rev. Sociol.* **27**: 415–44.
Merton, Robert K. (1957), *Social Theory and Social Structure*. Glencoe, IL: The Free Press.
Myers, David G. (1982), "Polarizing Effects of Social Interaction", pp. 125–61 in H. Brandstätter, J.H. Davis and G. Stocker-Kreichgauer (eds), *Group Decision Making*, London: Academic Press.

Michael Mäs - 9781789906851

Opp, Karl-Dieter (1984), “Balance Theory: Progress and Stagnation of a Social Psychological Theory”, *Philos. Soc. Sci.* **14** (1): 27–50.
Opp, Karl-Dieter (2001), “How Do Norms Emerge? An Outline of a Theory”, edited by R. Boudon, P. Demeulenaere and R. Viale, *Mind Soc.* **3** (2): 101–28.
Opp, Karl-Dieter (2013),“Rational Choice Theory, the Logic of Explanation, Middle-Range Theories and Analytical Sociology: A Reply to Gianluca Manzo and Petri Ylikoski Theory and Methods/Théorie et Méthodes”, *Soc. Sci. Inf.* **52** (3): 394–408.
Pradelski, B.S.R. and H.P. Young (2012), “Learning Efficient Nash Equilibria in Distributed Systems”, *Games Econ. Behav.* **75**: 882–97.
Richerson, Peter J. and Robert Boyd (2005), *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press
Rizzolatti, Giacomo and Laila Craighero (2004), “The mirror-neuron system”, *Annu. Rev. Neurosci.* **27** (1): 169–92.
Salzarulo, Laurent (2006), “A Continuous Opinion Dynamics Model Based on the Principle of Meta-Contrast”, *J. Artif. Soc. Soc. Simul.* **9** (1).
Schultz, P.W., J.M. Nolan, R.B. Cialdini, N.J. Goldstein and V. Griskevicius (2007), “The Constructive, Destructive and Reconstructive Power of Social Norms”, *Psychol. Sci.* **18** (5): 429–34.
Selten, R. and J. Apesteguia (2005), “Experimentally Observed Imitation and Cooperation in Price Competition on the Circle”, *Games Econ. Behav.* **51** (1): 171–92.
Sherif, M. (1935), “A Study of Some Social Factors in Perception”, *Arch. Psychol.* **27** (187): 1–60.
Smaldino, Paul E. and Joshua M. Epstein (2015), “Social Conformity Despite Individual Preferences for Distinctiveness”, *R. Soc. Open Sci.* **2** (3).
Snijders, Tom A.B., Gerhard G. van de Bunt and Christian E.G. Steglich (2010), “Introduction to Stochastic Actor-Based Models for Network Dynamics”, *Soc. Networks* **32** (1): 44–60.
Squazzoni, Flaminio (2012), *Agent-Based Computational Sociology*. Oxford: Wiley.
Steglich, Christian, Tom A.B. Snijders and Michael Pearson (2010), “Dynamic Networks and Behavior: Separating Selection from Influence”, *Sociol. Methodol.* **40** (1): 329–93.
Tajfel, Henri and John C. Turner (1979), “An Integrative Theory of Intergroup Conflict”, pp. 33–47, in W.G. Austin and S. Worchel (eds), *The Social Psychology of Intergroup Relations*, Monterey, CA: Brooks/Cole.
Tajfel, Henri and John C. Turner (1986), “The Social Identity Theory of Intergroup Behavior”, pp. 7–24 in S. Worchel and W.G. Austin (eds), *Psychology of Intergroup Relations*, Chicago, IL: Nelson-Hall Publishers.
Takács, K., Andreas Flache and M. Mäs (2016), “Discrepancy and Disliking Do Not Induce Negative Opinion Shifts”, *PLoS One* **11** (6).
Thaler, Richard H. and Cass R. Sunstein (2009), *Nudge: Improving Decisions About Health, Wealth, and Happiness*, New York: Penguin.
Urbig, Diemo and Robin Malitz (2005), “Dynamics of Structured Attitudes and Opinions”, pp. 206–12, in K.G. Troitzsch (ed.), *Representing Social Reality. Pre-Proceedings of the Thrid Conference of the European Social Simulation Association*, Koblenz.
Vignoles, V.L., X. Chryssochoou and G.M. Breakwell (2000), “The Distinctiveness Principle: Identity, Meaning, and the Bounds of Cultural Relativity”, *Personal. Soc. Psychol. Rev.* **4** (4): 337–54.
Vinokur, Amiram and Eugene Burnstein (1978), “Depolarization of Attitudes in Groups”, *J. Pers. Soc. Psychol.* **36** (8): 872–85.
Voss, Thomas (2001), “Game-Theoretical Perspectives on the Emergence of Social Norms”, pp. 105–36, in M. Hechter and K.-D. Opp (eds), *Social Norms*, New York: Russell Sage Foundation.
Ylikoski, Petri (2013), “The (Hopefully) Last Stand of the Covering-Law Theory: A Reply to Opp”, *Soc. Sci. Inf.* **52** (3): 383–93.
Young, H. Peyton (2001), *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Princeton, NJ: Princeton University Press.
Young, H. Peyton (2015), “The Evolution of Social Norms”, *Annu. Rev. Econom.* **7**: 359–87.

Michael Mäs - 9781789906851

11. Structure

Emily Erikson and Eric Feltham

Structure is an odd beast. For analytical sociology in particular, it poses interesting and difficult questions to address. Analytical sociology has been linked to versions of methodological individualism in which action, intentionality, and causal power originates with individuals (Boudon, 1971). Yet there is good reason to believe that structure has an independent effect on the world that does not lie within any one individual, but is instead located in the arrangement of individuals' relationships with one another. This possibility raises a series of conceptual questions. Does structure have causal power? If it is causal, is it an agent? Is the causal power of structure somehow different from agency?

1. PERSPECTIVES IN PHILOSOPHY

Whether structure, social or otherwise, genuinely exists remains controversial in the field of philosophy (Barnes, 2016; Rea, 1997; Haslanger, 2015). The controversy has long roots. The problem of structure is famously posed in *Parmenides*, where Socrates doubts that objects have a form, and instead posits that "… things are in fact just what we see …" (Plato, 2005: p. 607). Karl Popper's influential critique of the reality of *social* structure is particularly prominent; he accused traditional social theorists—notably Hegel—of "reifying" states. This objection has become a common refrain and has often been levied as a criticism indicating that researchers have not fully plumbed the complexities of state organization. Popper's view, however, had broader implications and drew directly on J.S. Mill, who took the phenomena of society to be "nothing but the laws of the actions and passions of human beings" (Epstein, 2015). Popper's objection was based on a denial of the existence of an independent reality above individuals, such as might come about through organization, roles structures, networks of relations, or rules of governance.

Popper's perspective has been shared by others. Kathrin Koslicki noted that twentieth-century metaphysical work has been dominated by what she terms the "standard mereology" paradigm (Koslicki, 2008), which she attributes to Stanislaw Leśniewski, Henry Leonard, and Nelson Goodman. This account does not distinguish "heaps" of objects with "structured wholes" of the same objects (Doepke, 1997; Koslicki, 2008). More precisely, standard mereology holds that there is no object besides the objects in the heap that differentiates the structured whole from the heap.

An example commonly used in the literature on the metaphysics of constitution helps illustrate the perspective. A sculptor crafts a lump of gold into a statue of Goliath. It may seem natural to believe that the statue possesses different properties than the lump of gold from which it is constructed. However, the gold still exists. So, one might ask whether there are now two objects: the lump of gold and the statue of Goliath. Is there anything over and above the lump of gold that yields the statue or are the statue and the lump of gold the same (Thomson, 1997)? The two objects are spatially coextensive. One response to this dilemma

Emily Erikson and Eric Feltham - 9781789906851

is the "one-many view" (Doepke, 1997). In this view, the lump of gold and the statue are considered different aspects of a single object. It follows that there may be many objects—the constituents—but the statue does not genuinely exist. Instead, it is merely a shorthand reference for a set of objects.

While this view has held in philosophy for some time, it has left some unsatisfied for several reasons. First, it seems counter-intuitive to deny that the statue exists. Second, the view does little to explain the new causal powers held by the statue versus the gold—to command a greater purchasing price, for example. It becomes even harder to ignore the effect of the arrangement of parts in the example of the difference between a car and a heap of automotive parts. Third, the standard mereology view has difficulty explaining how objects can persist over time even when its parts are replaced. If there is no object over and above the elements, how does a complex object continue to exist over time, even when its parts are replaced? For example, we might wish to state that the Boston Symphony Orchestra has played at Tanglewood each summer for eighty years, even though all of the musicians have been replaced over time.

Subsequently, recent work in analytic metaphysics has pushed back against the standard mereology paradigm. In *The Structure of Objects*, Kathrin Koslicki (2008) resurrects Aristotle's distinction between *content* and *form* to create a new metaphysics of the composition of objects. Aristotle's framework had been rejected due to difficulties in defining the nature of *form*. In Koslicki's framework, form can be defined as a structure. Structure is further defined as an entity that "make[s] available positions or places for other objects to occupy". For her, structures themselves are taken to be objects rather than properties or relations *of objects*. This is to say that the set of relations between the constituents is a property of an overarching complex object. She further adds that any particular occupant of a slot in the structure must satisfy general type restrictions set for the slot, i.e. only particular occupants satisfying these type constraints are admissible. Armed with these definitions one can conclude that under admissible transformations, structures can persist beyond the replacement of parts, e.g. the Boston Symphony Orchestra remains the Boston Symphony Orchestra despite a complete turnover in musicians from 1937 to 2019.

Philosophers have also conceptualized *social* structure. Building on Koslicki and Haslanger, Katherine Ritchie (2020) defines social structures as "networks of relations that are constitutively dependent on social factors" (p. 407). This definition is not identical to social networks, however, as her networks can include nodes that are not exclusively persons or firms. Additionally, her definition of social structure is based on constitutive rather than causal dependence. Constitutive dependence is stronger than causal dependence because elements that are constitutively dependent are not conceivable as independent entities (Haslanger, 1995; Ylikoski, 2013). Structure S then may be even more precisely defined as "social" when at least one of the following conditions hold: "(i) in defining what it is to be S, reference must be made to some social factors, (ii) social factors are metaphysically necessary for S to exist or (iii) social factors 'ground' the existence of S (or the fact that S exists)" (Ritchie, 2020). As an example, the Boston Symphony is constituted by the organized pattern of relationships between the conductor, strings, woodwinds, brass and percussion. The orchestra cannot exist without that pattern of relations, thus there is a constitutive relation. Whether they sound good or not may be related to that organizational structure, but the relationship between performance and organization, if there is one, would be causal not constitutive.

The account of social structure outlined in these works is agnostic with regard to *intention*. Many social structures were not intentionally created or planned, yet are constitutively

Emily Erikson and Eric Feltham - 9781789906851

dependent on the social. Favelas (informal settlements), traffic rules that emerge without legal stipulation, or any emergent phenomenon, are examples that apply here.

2. PERSPECTIVES FROM THE SOCIAL AND NETWORK SCIENCES

Social scientists have often employed definitions of social structure that are extremely Catholic, where social structures are the result of any human action and stand only in contrast to naturally occurring phenomena. Drawing from this tradition, Haslanger (2015) defines social structure loosely as institutions, practices, conventions, hierarchies and other social entities. However, while loose definitions serve as a conceptual starting point, they are often so broad as to give little insight. Mathematics defines structure more precisely as: (i) an ordered set of objects, (ii) a list of relations between the objects, and (iii) the properties required by the relations (Koslicki, 2008; Mac Lane, 1996). Through the incorporation of graph theory into the study of social relations, a similar definition of structure has become influential in sociology, where structure is conceived as the order of relations between social actors. Network-oriented accounts of social structures became increasingly prevalent in the late-twentieth century, and their numbers expanded at an ever greater pace in the twenty-first. Since any attempt to review structure in the more encompassing sense would potentially include all of sociological inquiry, we restrict ourselves to the idea of structure formalized in graph theory and mobilized in the study of social relations by social network analysis.

In the early decades of the development of the field of social network research, the structure of social relationships between people was a central area of inquiry. Block modeling in particular was a tool that held out the promise of unlocking the importance of social structure through a systematic, empirically based computational method for describing it. In pioneering work, White et al. (1976) first redefined social structure as patterns of relations rather than norms and values or group attributes (p. 733). They then provided a method for systematically describing different types of patterns of relations, i.e. a means through which to categorize network structures. These structures, the researchers argue, define role sets occupied by individuals, as well as the relations between role sets, which define the regularities of social existence and give different structures distinct characteristics—which could be recognized, for example, as a propensity for conflict or cooperation, efficiency, or hierarchical rigidity. Other researchers later used block modeling to create distinct subtypes, such as cohesive subgroups, center-periphery, centralized, transitive, or hierarchical, to categorize network types (Faust and Wasserman, 1992). Though they were not then able to realize it in this early work, White et al. defined their ultimate goal to be "to show how concrete social processes and individual manipulations shape and are shaped by structure" (p. 773).

Paul Erdös and Alfréd Rényi (1960) developed a random graph model that could be applied to social network analysis. The random graph is chosen either at random from the set of all possible graphs for some set of nodes and arcs or is constructed by adding arcs at random between a set of defined nodes. The properties of Erdös–Rényi random graphs became fairly well understood, but the usefulness of applying random models to a structured world of social connections has often been contested. Early small-world researchers attempted to theorize, describe and measure the way that different patterns of social ties could increase

Emily Erikson and Eric Feltham - 9781789906851

or decrease social distances between groups and individuals within those groups (Travers and Milgram, 1969; Hunter and Shotland, 1974; Shotland, 1976; Stoneham, 1977; Bernard and Killworth, 1978). Percolation models were developed that could be applied to social interactions (Shante and Kirkpatrick, 1971) and new techniques of multi-dimensional scaling and factor analysis were used to describe systems and patterns of relations (Shepard and Arabie, 1979). Additionally, a number of largely continuous measures were developed or imported into social network analysis. Examples include density, which measures the number of ties in a network as a proportion of the possible number of ties, giving a rough idea of the intensity of social interactions in a network. Centralization calculates, through several different means, the extent to which one or a small number of nodes hold more central positions in a network, i.e. link together many other nodes. Clustering, a larger and more complex suite of methods, uses various means to identify sub-regions within networks that are more densely interconnected. All of these measures have been linked through empirical research to diffusion and contagion processes and a variety of social outcomes.

In the 1990s, social network analysis turned toward a more positional style of analysis. Analysts sought to investigate the ways in which an individual's location within a structure could impact their outcomes. The field flourished, but not through an analysis of the independent effect of structure. The positional turn is perhaps best exemplified in the difference between Mark Granovetter's "The Strength of Weak Ties" (1973) and Ron Burt's *Structural Holes* (1992). Granovetter's influential article built a theory on elementary principles and assumptions about patterns of interaction that demonstrated the importance of relationships with people outside the close family and circles of friends. Granovetter's theory presumes that these relationships will be weaker in that individuals necessarily must spend less time on them. Since these relationships are not components of a cohesive cluster of redundant ties (as in family or dense social cliques), but instead span across clusters, they are important conduits of new and valuable information and resources. In formalizing the theory behind this insight, Granovetter described the larger structural conditions that could produce weak ties; he also paid a significant amount of attention to the difference in outcomes for communities that possessed weak ties versus those that did not. At heart, his unit of analysis was the network structure of the group.

In the 1990s, Ron Burt (1992) published his book, *Structural Holes*, which built on Burt's (1982) previous research, his long-term debate with James Coleman's idea of social capital, and Granovetter's theory of weak ties. Burt's emphasis differed significantly from Granovetter's in that he refocused the problem of network structure on individual outcomes, making it more positional in nature. He formalized a local measure of individual structural constraint that gauges the potential for individuals to successfully leverage their network position into a competitive advantage in various market-like arenas. In Burt's analysis, network structure is not the primary unit of analysis. It is instead the individual. The position of the individual in the network mediates their behavior and outcomes. Burt's contribution presaged a decade-or-so long period in which much of social network analysis was focused on analyzing the impact of network position on individual and firm performance. Part of the shift from macro-structure to positional analysis may have been related to the difficulties involved in collecting and analyzing enough data to completely represent and compare whole networks. It is much easier to collect and analyze the position of many individuals in a single network than to collect and analyze many different networks, which are of course always composed of numerous individuals or agents of some kind.

Emily Erikson and Eric Feltham - 9781789906851

By the end of the 1990s, a resurgence of attention to the macro-structure of social networks had begun to take shape. In a return to the small-worlds literature, Duncan Watts and Steve Strogatz formalized a definition of what type of structural configuration should be considered a "small world." Small-world networks are defined by levels of local clustering much higher than in random graph models, but with comparably low average path lengths (Watts and Strogatz, 1998; Watts 1999). The two coauthors showed that small-world network structures are mathematically likely, in that there are a broad range of networks that fall into the small-world category by virtue of the greater susceptibility of path length than clustering to small shifts in the larger network structure. And importantly, they convincingly demonstrated that the small-world structure has important implications for diffusion and collective dynamics. In so doing, they created a new conceptual category through which entire network topologies may be evaluated.

At roughly the same time, Albert-László Barabási and Réka Albert (1999) developed a different model: the scale-free network. Scale-free networks are defined by the degree distribution of nodes following a power law. These networks tend to have a tree-like appearance. In such a structure, the average is not a particularly descriptive statistic because of the extreme dispersion in nodal degree at the tail end of the distribution. Similar to the small-world network, this type of network structure captures some basic elements of human interactions. In this case, it is a Matthew effect, where the possession of riches confer even greater riches (Merton 1968). In social interactions, the idea is that each additional friend a person has increases the likelihood of making a new friend. This dynamic seems like a reasonable approximation of interaction likelihoods in places like Twitter, where we observe an extreme dispersion at the tails of the distribution of accounts and followers: most people have few followers, but a small number of individuals have a very large number of followers. It has since been found that truly scale-free social networks are rare (Broido and Clauset, 2019). However, a broader class of preferential-attachment networks, including social networks that are weakly scale-free, has proved to be another useful category through which to explore network topology.

These two theoretical models, along with the random network, came to represent the three main categories researchers most frequently use to describe and compare social network structures. Regular ring lattice networks are used to represent clustered networks, and their properties have been theorized in the literature on complex contagions (Centola and Macy, 2007; Centola, 2018). There are also several highly structured models, such as ring networks, star networks, and tree networks which provide useful conceptual tools. And it has been argued that complex financial networks should sometimes be described by a mixture of different network typologies, e.g. clustered, small-world, and preferential attachment, that occupy different sub-areas in the larger population (Mani and Moody, 2014).

With the development of these models, it became common to test the impact of structure on group outcomes. The implications in the differences between small-world networks and preferential attachment networks have been explored at great length for contagions and diffusion processes (Watts, 1999; 2004). Brian Uzzi and Jarrett Spiro (2005) explored the impact of small-world structure for creativity in Broadway musicals. Michael Kearns, Siddarth Suri, and Nick Montfort, for example, found individuals linked together in a small-world formation perform better than individuals in preferential attachment networks when brought together to solve collective problems in a decentralized decision process (Kearns et al., 2006). These examples are but a few in the vast relevant literature.

Emily Erikson and Eric Feltham - 9781789906851

In recent promising extensions, researchers have been able to apply experimental methods to the study of social networks by using structure as a manipulable treatment in online settings. For example, Damon Centola (2010) recruited individuals into an online platform used for disseminating health information. Participants were assigned "health buddies" to whom they were connected and with whom they were meant to interact in order to reinforce positive health behaviors. Unbeknownst to the participants, the larger structure of the network connecting individuals was manipulated into two different patterns: a random network and a ring lattice network. Individuals were randomly placed into the different types of network structures. The differences in health behavior were then measured and the rate of adoption was compared across the different network topologies. The result was a higher proportion of adoption in the lattice-type network. The great advantage of this approach is the ability to randomize across individual characteristics and to hold constant and control for different aspects of the local structural environment, such as average degree, in way that is rarely possible in face-to-face or offline interactions. Thus, the experiment provides strong evidence of the causal impact of structure in and of itself. The online approach has also been usefully reviewed by Centola (2018) and Salganik (2018), and an entire software suite, Breadboard, has been developed to facilitate future online network experiments (McKnight and Christakis, 2016).

Along with the new categorical vocabulary, a renaissance in clustering methods began in the early 2000s. Mark Newman, a physicist, in particular contributed several important elegant new methods for clustering networks (Clauset et al., 2004; Newman, 2004; Newman, 2006). Much of this work was fueled by an increased interest in network problems in the hard sciences (Watts, 2004). Despite these rapid advances, there remains suggestive evidence that other, less well-studied measures of network topology may also have significant impacts on social processes. For example, it has recently been shown that the chromatic polynomial can have a significant impact on coordination in networks. The chromatic polynomial was designed by George David Birkhoff and D.C. Lewis (1946) to explore solutions to the graph coloring problem. Michael Kearns et al. (2006) reconceived the graph coloring problem as a networked coordination game in which the goal is for each node to choose a color different from its neighbors' colors. The chromatic number is the lowest number of colors needed to satisfy the goal of the game; each node is colored differently from its neighbors. The chromatic polynomial is a function that produces the number of possible solutions to the problem in a given network for a given number of colors. In practice, it is used to refer to the number of possible solutions for the graph's chromatic number, i.e. the number of solutions for the minimal number of colors necessary for the problem to have a solution. It is intuitive to suppose that the number of solutions would have an impact on the success rate for the game, and recent research suggests that this is indeed the case (Erikson and Shirado, 2020). What is particularly interesting about the chromatic polynomial is that it is difficult to observe, not well correlated with other network measures, highly susceptible to small changes, has an unusually high variance, and may have a significant impact on important social outcomes.

3. PERSPECTIVES IN ANALYTICAL SOCIOLOGY

The way in which analytical sociology has treated structure has evolved over time. In the 1998 book, *Social Mechanisms: An Analytical Approach to Social Theory* (Hedström and Swedberg, 1998), structure is described as a mechanism. In particular Hedström and Swedberg point to

Emily Erikson and Eric Feltham - 9781789906851

network diffusion as a mechanism that links information flow to adoption, using the example of physicians and drug prescriptions. In contrast to some of the philosophical perspectives previously presented, their explanation indicates that they conceptualize structure as a relationship between objects, rather than as an object itself. In *Dissecting the Social*, Hedström (2005) describes network processes as mechanisms that connect micro- and macro-level outcomes. Hedström argues that the network mechanisms always operate through the individual, whose decisions to act ultimately produce the outcome of interest. In his words, it is "individuals, not social entities that are endowed with causal powers" (p. 115).

In *The Handbook of Analytical Sociology*, Hedström and Bearman (2011) lay out a more detailed understanding of how network structure fits into the theoretical architecture of analytical sociology. Like Ritchie, their understanding of the relationship between structures and occupants is not defined by a causal relationship. Hedström and Bearman characterize it as a supervenience relation. Supervenience is a relationship in which for any two sets of properties A and B, A supervenes on B if and only if there cannot be change in the A-properties without a change in the B-properties (McLaughlin and Bennett, 2018). This relationship is asymmetric: e.g., the musicians in the Boston Philharmonic may change without the Philharmonic itself changing properties. However, the Philharmonic changing its properties implies that the musicians also change. Here, A and B respectively refer to the macro- and micro-levels of social phenomenon. This means that any change in macro-structure implies a change in—or is indeed equivalent to—a change in the micro-level, but the micro-level can change without changing the macro-level. This asymmetry allows us to call the structure, in this case a symphony orchestra, a stable relation that is composed only of its parts and without causal powers independent from its parts. In this view, structural processes are accurately described as processes in which changes in macro-structure entail changes in micro-structure, and it is the micro-structural changes which effect differences in social outcomes. Structure only works through its impact on individuals, who respond to the structural condition they inhabit. Only through the actions of individuals does structure have an impact on the world.

In more recent work, Gianluca Manzo (2014) moved to broaden the role of structure in analytical sociology. He recommends that sound analytical explanation will always include a specification of both entities and "a realistic description of the structural interdependences that are likely to relate these entities" (p. 18). And, in an important change Manzo construes the relationship between micro and macro layers as one of dynamic co-determination. He is also careful to note that structure, and all macro-level properties, are potentially emergent and therefore more than simply the sum of their parts. In the next sections we will try to put these different perspectives into some dialogue with each other to shed light on a series of questions about the nature of social structure.

4. IS SOCIAL NETWORK STRUCTURE AN OBJECT?

It seems uncontroversial to say that physical objects exist in the world. However, an arrangement of objects does not intuitively satisfy the definition. If social networks are arrangements of relations between social actors, then are they a different sort of entity from objects or a special kind of object? And if they are not objects, can they be causal actors? The analytical sociological perspective seems to concur with the position that the arrangement of objects is not an object, which indeed has some prima facie validity. Ritchie and the new challenges to

Emily Erikson and Eric Feltham - 9781789906851

standard mereology in philosophy, however, appear to provide metaphysical grounds through which to think about social structures, understood through social network topology as objects in themselves. In this view, however, the objects are bundles of other objects and relations, and relations are not objects. At this point, it becomes hard to know how to define what is or what is not an object. Does treating social networks as objects reify dynamic processes? Where does something like a social norm fit in?

Empirical research has treated social networks as objects. There is, for example, a flourishing literature on networks of networks (NoN), which considers interdependencies between different networked systems, such as transportation networks, power networks, and telecommunication networks. While these are often analyzed as interconnected layers, the networks are also usefully represented as nodes in a larger network of networks (Gao et al., 2014; D'Agostino et al., 2014). In organizations research there is similarly a robust stream of research on inter-organizational networks, which necessarily link across the intra-organizational networks contained within their target firms (Gulati and Gargiulo, 1999; Baker and Faulkner, 2017).

This conceptualization of networks as objects or nodes is supported in graph theory. In math, for example, structures may be elements of larger structures through recursion. This nesting of structures in larger structures hints at the treatment of structure as an object, rather than a property or relation. As such, it appears to be consistent with Ritchie's position that structure is an object, albeit one subject to fairly stringent definitional criteria.

5. WHAT IS THE RELATIONSHIP BETWEEN THE SOCIAL AND SOCIAL NETWORKS?

There seems to be some consensus across positions that the relationship between the social and social structure is even more binding than simple cause and effect. Ritchie (2020) argues that social structure is constituted by the social. Hedström and Bearman (2009) argue that the micro-layer of human existence, individual activity, and the layer of the structure of interactions are related by supervenience, in which the layers are so closely linked that a change in one implies a change in another. In this sense, the supervenience relationship also seems to be a constitutive dependency. However, Richie is more concerned with identifying what forms of structure should be considered social, whereas Hedström and Bearman seem to be more motivated by careful consideration of whether the relationship between social structure and individual behavior is distinct from one of cause and effect.

Richie's claim can be considered less binding than Hedström and Bearman. For example, her claim might in fact be made stronger if one posited that all social phenomenon must have both content and structure. Can people interact in an unstructured manner—or is this always itself a type of structure? We tend to associate social environments with a lack of rigid social roles with a lack of structure; however, these are often undergirded by highly structured formal rules. For example, a crowd may be considered an unstructured group; yet social theorists have convincingly argued that these manners of association possess specific underlying structural properties (Le Bon, 1925; Granovetter, 1978). Even groups without interconnections have a certain structural character. Consider Marx's (2001) sack of potatoes or the masses in Arendt's (1968) *The Origins of Totalitarianism*. Is it possible for individuals to operate without structure? Even isolation may be conceived as a structural condition. Social existence without social

Emily Erikson and Eric Feltham - 9781789906851

structure may be even more difficult to conceive than existence without space and time. Perhaps social structure and social actors are dual in the sense that both must exist for one to exist.

This claim, however, is different than the even stronger description of the relationship as one of supervenience. Does a change in the macro-structural level always imply a change in the micro-level? The answer might vary depending on whether the micro-level is defined as individuals or relationships. There are different positions on this in sociology. Relationalists have argued that the relevant micro-layer of social existence is relationships or interactions between people (Erikson, 2013). If relationships are the micro-layer, it does seem true that a macro-layer that designates an ordering of relations between individuals must necessarily change if even one of the relationships that make it up changes. However, if there is redundancy in the network, changes in one or two relationships may not change the outcome of different processes on the network.

But if analytical sociology's micro-foundation is individuals, we should ask if a change in macro-structure necessarily changes the individuals who occupy that social structure. This strong relationship does not seem to hold in philosophers' conceptualization of structures more generally. For example, the gold that makes up the Goliath statue could also be made into a statue of Gertrude Stein. The gold does not change. Similarly, taking the wheels off a car does not change any of the individual pieces of the car—it just interrupts a central connection crucial to the car's functioning. Is the relationship between social structure and individuals more binding?

Another way of thinking through this question in more detail is to think about the ways in which changes in structure may produce concurrent changes in individuals. Within the analytical sociology framework, structure can influence individual behavior through desires, beliefs, and opportunities. We might ask, is a change in opportunity a change in the individual? What if the individual does not observe the opportunity? One problem with the idea that structure only channels individual behavior is that the complexities of structure can be difficult to observe—for example, the fact that community structures impact the spread of viral contagion is not something that is consciously observed or understood by most members of a community. Yet it affects them. Indeed, it seems evident that structure matters whether individuals are consciously aware of it or not (Podolny and Lynn, 2009). But if changes in structure cause changes in individuals whether or not they are consciously aware of it, the contradiction might be resolved. If structure constitutes individuals, it must be so, but not all would necessarily agree with this proposition. The onus for proving this lies with sociologists, and we might at least agree that overwhelmingly conclusive evidence has not yet been accumulated.

On the other hand, several theorists take the position that all social phenomena reduce to psychological states (Dretske, 1988; Jackson and Pettit, 1992). Jackson and Pettit, for example, argue that it is unclear that "most of the factors invoked in structural explanations" have any other pathway to affect human behavior and belief other than through (i) coming into their consciousness and explicitly driving behavior, or (ii) "covertly" altering out behavior, unbeknownst to the actor.

6. ARE SOCIAL NETWORKS CAUSAL ACTORS?

Another way of trying to disentangle these problems is to think through the problem of whether structure can have a causal effect on actors. Cause and effect is a relationship that is different from the relationships of constitutive dependence and supervenience. If one thing constitutes

Emily Erikson and Eric Feltham - 9781789906851

another, they might be considered two aspects of the same phenomenon. A relationship of cause and effect implies the cause is an external force, like the cue ball that hits the eight in a game of pool. If structure can have a causal effect, this could imply that it is distinct from the social phenomenon that it effects. It also implies by definition that structure is an agent, since agents are entities that are able to produce effects.

It is on the question of the causal efficacy of networks that the research in network analysis has weighed in most heavily. In at least the last three decades of social network research, the central question that has been addressed is whether networks are causal. The answer within the field is clearly a resounding yes. However, the implications for positional and structural approaches to network causality may be different.

As described earlier, positional research analyzes the impact of the position occupied by an individual within a network on their behavior, for example through the information that is presented to them or the options they have open to them. These works treat networks as pipes and prisms that channel individual behavior (Podolny, 2001). Hedström and Bearman's perspective emphasizes that channeling is different from causing. Water that irrigates a field may be directed there by a canal, but the canal does not cause the plants to grow. The canal cannot cause growth without water, but water can cause growth without the canal. Translated into social network analysis, network position is not a cause acting on individuals. It is a channel through which other causes flow. Hedström (2005) has particularly emphasized that a position within the network is not an agent because it cannot produce an effect without acting through the medium of a person. Network position works by influencing the state of the individual, who then acts under its influence. This situation is why the relationship is one of supervenience. The assumption is that network position changes the person, who then initiates a causal sequence. Thus, while not disagreeing with the importance of network analysis, Hedström and Bearman can still discount its role as an agent.

It is difficult to challenge this perspective within the framework of positional network analysis. This is because, in this research, the unit of analysis is the individual or firm and network position is a covariate. Research that looks at the structure of the network as a whole provides a more formidable challenge to the idea that structure is not an agent.

The small-world network theory can be used to think around an agent-based perspective, by which we mean an analytical framework that assumes that agents have causal power and structures merely mediate that power. It is well documented that a small-world structure has significant implications for contagion rates and the diffusion of information through a network (Watts 1999). The question then is whether the causal effect comes from the individuals who make up the nodes or the structure itself, i.e. the patterned interaction between nodes. The position that the structure does not have causal power could manifest itself in two different ways. First, the structure merely imposes constraints on the actions of individuals. Second, a slightly weaker but still possible explanation would be that the structure itself is the product of individual decisions and actions. In either case, the structure is just a mediating factor, not a causal factor.

The problem with the first position is that average shortest path length, which is so crucial to the diffusion properties of small-world networks, is a global network measure. Path length does not depend on the local environment of nodes. For example, two networks with the same degree distribution could have very different average path lengths depending on how the network was arranged. This possibility suggests that people can be situated within the same local structural conditions, but experience different outcomes depending on the global

Emily Erikson and Eric Feltham - 9781789906851

structure. In such a case, it becomes difficult to maintain that the structure itself, outside of the way it impacts the experience of each individual within the network, is not having an independent effect on outcomes. The second position is also potentially untenable with respect to small-world networks. One of the most elegant properties of the small-world theory is the way in which the authors show that the introduction of only a very small amount of randomness dramatically changes average path length. While it is true that individuals do on occasion purposively take random action, it is also true that randomness connotes an absence of purpose. Indeed, randomness may be most likely to occur when purpose is interrupted. And the large-scale impact of the introduction of randomness in clustered networks is the result of the mathematical properties of networks. Thus, again it seems at least potentially problematic to say that individual decisions create small-world networks. It seems just as likely that individuals' inability to fully control their actions creates small-world networks given the mathematical properties of graphs.

This position is fleshed out in Duncan Watts and Peter Dodds's (2007) work on accidental influentials. Revising decades of research that attempted to identify the most influential individuals in a community, Watts and Dodds showed that no one individual impacts diffusion rates nearly as much as the shape of the total distribution of nodal degree in networks. Influential people are only influential in certain network structures. In these cases, individuals can be affected by structural patterns that may be both small and outside of their sphere of observation and experience. It becomes more difficult in these cases to argue that structure always acts through influencing an individual. Interestingly, however, Manzo et al. (2018) found that a negative signal from a structurally influential individual can alter the spread of innovation. Thus, more research may yet tell a different tale, as we accumulate evidence that will allow us to think even more clearly about the impact of structure, and particularly network topology, on social processes and outcomes.

CONCLUSION

Social structure can be an amorphous term in the social sciences. We have purposefully taken a narrow approach to defining it here as social network topology. This focus allowed us to explore some of the ways that that work could be useful for thinking through social processes—and also how knowledge of structure in the social sciences could be useful for philosophers thinking in more abstract terms. While more empirical work is always helpful, theoretical work will also be necessary to continue to improve the conceptual and methodological tools available to social science researchers.

REFERENCES

Arendt, Hannah (1968), *The Origins of Totalitarianism*, New York: Harcourt, Brace, Jovanovich.

Baker, W.E. and R.R. Faulkner, (2017), "Interorganizational networks," in *The Blackwell Companion to Organizations*, Oxford, UK: Blackwell, 520–40.

Barabási, A.L. and R. Albert (1999), "Emergence of scaling in random networks", *Science*, **286** (5439): 509–12. doi:10.1126/science.286.5439.509

Barnes, E. (2016), "Realism and social structure", *Philosophical Studies*, **174** (10): 2417–33. doi:10.1007/s11098-016-0743-y

Bernard, H.R. and P.D. Killworth, (1978). *A Review of the Small World Literature*. Report BK-115-78, West Virginia University, Morgantown Department of Sociology, July.

Birkhoff, G.D. and D.C. Lewis (1946), "Chromatic Polynomials", *Transactions of the American Mathematical Society*, **60** (3). doi:10.2307/1990348
Bon, G. Le (1925), *The Crowd: A Study of the Popular Mind*, New York: MacMillan.
Boudon, Raymond (1971), *The Uses of Structuralism*, London: Heinemann.
Broido, A.D. and A. Clauset (2019), "Scale-free networks are rare", *Nat Commun.*, **10** (1): 1017. doi:10.1038/s41467-019-08746-5
Burt, R.S. (1982), *Toward a Structural Theory of Action*, San Diego, CA: Academic Press.
Burt, R.S. (1992), *Structural Holes: The Social Structure of Competition*, Cambridge, MA: Harvard University Press.
Centola, D. (2010), "The spread of behavior in an online social network experiment", *Science*, **329** (5996): 1194–7. doi:10.1126/science.1185231
Centola, D. (2018), *How Behavior Spreads: The Science of Complex Contagions* (Vol. 3), Princeton, NJ: Princeton University Press.
Centola, D. and M. Macy (2007), "Complex contagions and the weakness of long ties", *American Journal of Sociology*, **113** (3): 702–34. doi:10.1086/521848
Clauset, A., M.E. Newman and C. Moore (2004), "Finding community structure in very large networks", *Phys Rev E Stat Nonlin Soft Matter Phys*., **70** (6 Pt 2), 066111. doi:10.1103/PhysRevE.70.066111
D'Agostino, G. and A. Scala (2014), *Networks of Networks: The Last Frontier of Complexity*, New York: Springer.
Doepke, F. (1997), "Spatially Coinciding Objects", in M. Rea (ed.) *Material Constitution: A Reader*, Lanham, MD: Rowman and Littlefield, xv–lvii.
Dretske, F. (1988), *Explaining Behavior: Reasons in a World of Causes* (Vol. 100), Cambridge, MA: MIT Press.
Epstein, B. (2015), *The Ant Trap: Rebuilding the Foundations of the Social Sciences*, Oxford, UK: Oxford University Press.
Erdős, P. and A. Rényi (1960), "On the evolution of random graphs", *Publ. Math. Inst. Hung. Acad. Sc.*, **5** (1): 17–60.
Erikson, E. (2013), "Formalist and relationalist theory in social network analysis", *Sociological Theory*, **31** (3): 219–42. doi:10.1177/0735275113501998
Erikson, E. and H. Shirado (2020), "Network Structure, Property, and the Division of Labor", *American Sociological Review*, **86** (1): 758–86.
Faust, K. and S. Wasserman (1992), "Blockmodels: Interpretation and evaluation", *Social Networks*, **14** (1–2): 5–61. doi:10.1016/0378-8733(92)90013-w
Gao, J., D. Li and S. Havlin (2014), "From a single network to a network of networks", *National Science Review*, **1** (3): 346–56. doi:10.1093/nsr/nwu020
Granovetter, M. (1973), "The strength of weak ties", *American Journal of Sociology*, **78** (6): 1360–80. www.jstor.org/stable/2776392
Granovetter, M. (1978), "Threshold models of collective behavior", *American Journal of Sociology*, **83** (6): 1420–43. doi:10.1086/226707
Gulati, R. and M. Gargiulo (1999), "Where do interorganizational networks come from?", *American Journal of Sociology*, **104** (5): 1439–93. doi:10.1086/210179
Haslanger, S. (1995), "Ontology and social construction", *Philosophical Topics*, **23** (2): 95–125. www.jstor.org/stable/43154209
Haslanger, S. (2015), "What is a (social) structural explanation?", *Philosophical Studies*, **173** (1): 113–30. doi:10.1007/s11098-014-0434-5
Hedström, P. (2005), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge, UK: Cambridge University Press.
Hedström, P. and P. Bearman (2011), *The Oxford Handbook of Analytical Sociology*, Oxford: OUP.
Hedström, P. and R. Swedberg (1998), *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge, UK: Cambridge University Press.
Hunter, J.E. and R.L. Shotland (1974), "Treating data collected by the 'Small World' method as a Markov process", *Social Forces*, **52** (3): 321–32. doi:10.2307/2576887
Jackson, F. and P. Pettit (1992), "Structural Explanation in Social Theory", in K. Lennon and D. Charles (eds), *Reduction, Explanation, and Realism*, Oxford, UK: OUP.
Kearns, M., S. Suri and N. Montfort (2006), "An experimental study of the coloring problem on human subject networks", *Science*, **313** (5788), 824–7. doi:10.1126/science.1127207
Koslicki, K. (2008), *The Structure of Objects*, Oxford: Oxford University Press.
Mac Lane, S. (1996), "Structure in Mathematics", *Philosophia Mathematica*, **4** (2): 174–83. doi:10.1093/philmat/4.2.174
Mani, D. and J. Moody (2014), "Moving beyond stylized economic network models: The hybrid world of the Indian firm ownership network, *AJS*, **119** (8): 1629–69. doi:10.1086/676040
Manzo, G. (2014a), "Data, generative models, and mechanisms: More on the principles of analytical sociology", in G. Manzo (ed.), *Analytical Sociology*, Chichester, UK: Wiley, 1–52.

Emily Erikson and Eric Feltham - 9781789906851

Manzo, G. (2014b), *Analytical Sociology: Actions and Networks*, Chichester, UK: Wiley.
Manzo, Gianluca, Simone Gabbriellini, Valentine Roux, and Freda Nkirote M'Mbogori (2018), "Complex contagions and the Diffusion of innovations: Evidence from a small-N study", *Journal of Archaeological Method and Theory*, **25** (4): 1109–54. https://doi.org/10.1007/s10816-018-9393-z.
Marx, Karl (1852), *The 18th Brumaire of Louis Bonaparte*, London: Electric Book Co. c2001.
McKnight, M.E. and N.A. Christakis (2016), "Breadboard" computer software. *Breadboard: Software for Online Social Experiments*. Vers. 2. Yale University. Web. https://breadboard.yale.edu/
McLaughlin, B. and K. Bennett (Winter 2018), "Supervenience", in E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, Palo Alto, CA: Metaphysics Research Lab, Stanford University.
Merton, R.K. (1968), "The Matthew effect in science: The reward and communication systems of science are considered", *Science*, **159** (3810): 56–63. doi:10.1126/science.159.3810.56
Newman, M.E. (2006), "Finding community structure in networks using the eigenvectors of matrices", *Phys Rev E Stat Nonlin Soft Matter Phys.*, **74** (3 Pt 2), 036104. doi:10.1103/PhysRevE.74.036104
Newman, M.E. and M. Girvan (2004), "Finding and evaluating community structure in networks", *Phys Rev E Stat Nonlin Soft Matter Phys.*, **69** (2 Pt 2), 026113. doi:10.1103/PhysRevE.69.026113
Plato (2005), "Parmenides (127a-135d)", in S.M. Cohen, P. Curd, and C.D.C. Reeve (eds), *Readings in Ancient Greek Philosophy: From Thales to Aristotle* (3rd ed.), Indianapolis, IN: Hackett.
Podolny, J. and F. Lynn (2009), "Status", in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, New York: Oxford University Press.
Podolny, Joel M. (2001), "Networks as the pipes and prisms of the market", *American Journal of Sociology*, **107** (1): 33–60. doi:10.1086/323038
Rea, M. (1997), "Introduction", in M. Rea (ed.), *Material Constitution: A Reader*, Lanham, MD: Rowman and Littlefield, xv–lvii.
Ritchie, K. (2020), "Social structures and the ontology of social groups", *Philosophy and Phenomenological Research*, **100** (2): 402–24. doi:10.1111/phpr.12555
Salganik, M.J. (2018), *Bit by Bit: Social Research in the Digital Age*, Princeton, NJ: Princeton University Press.
Shante, V.K.S. and S. Kirkpatrick (1971), "An introduction to percolation theory", *Advances in Physics*, **20** (85), 325–57. doi:10.1080/00018737100101261
Shepard, R.N. and P. Arabie (1979), "Additive clustering: Representation of similarities as combinations of discrete overlapping properties", *Psychological Review*, **86** (2): 87–123. doi:10.1037/0033-295x.86.2.87
Shotland, R. Lance (1976), *University Communication Networks: The Small World Method*, New York: Wiley.
Stoneham, A.K.M. (1977), "The small-world problem in a spatial context", *Environment and Planning A: Economy and Space*, **9** (2): 185–95. doi:10.1068/a090185
Thomson, J.J. (1997), "Parthood and identity across time", in M. Rea (ed.), *Material Constitution: A Reader*, Lanham, MD: Rowman and Littlefield, xv–lvii.
Travers, Jeffrey and Stanley Milgram (1969), "An experimental study of the small world problem", *Sociometry*, **32** (4): 425. https://doi.org/10.2307/2786545
Uzzi, B. and Spiro, J. (2005), "Collaboration and creativity: The small world problem", *American Journal of Sociology*, **111** (2): 447–504. doi:10.1086/432782
Watts, D.J. (1999), "Networks, dynamics, and the small-world phenomenon", *American Journal of Sociology*, **105** (2): 493–527. doi:10.1086/210318
Watts, D.J. (2004), "The 'new' science of networks", *Annual Review of Sociology*, **30** (1): 243–70. doi:10.1146/annurev.soc.30.020404.104342
Watts, D.J. and P.S. Dodds (2007), "Influentials, networks, and public opinion formation", *Journal of Consumer Research*, **34** (4): 441–58. doi:10.1086/518527
Watts, D.J. and S.H. Strogatz (1998), "Collective dynamics of 'small-world' networks", *Nature*, **393** (6684): 440–2. doi:10.1038/30918
White, H.C., S.A. Boorman and R.L. Breiger (1976), "Social Structure from multiple networks. I. Blockmodels of roles and positions", *American Journal of Sociology*, **81** (4): 730–80. doi:10.1086/226141
Ylikoski, P. (2013), "Causal and constitutive explanation compared", *Erkenntnis*, **78** (S2): 277–97. doi:10.1007/s10670-013-9513-9

Emily Erikson and Eric Feltham - 9781789906851

12. Norms

Pierre Demeulenaere

Norms are a vast topic of study in sociological literature, maybe the most central one, since social life is largely organized through norms that are pervasive in all its dimensions. It is hard to find a domain of social life that is not concerned by social norms: food habits, architecture, shape of towns, etc. are, for instance, ruled by social norms that define typical choices in specific settings. Emile Durkheim had famously highlighted this in his classical writings. Sociological literature has since been divided into two main interpretive approaches: one that centres on individual interests and derives social norms from those (Hechter and Opp, 2005), and another that stresses "social" aspects and individuals' submission to norms they have not personally chosen. An analytical approach to social norms should not reproduce this widely held opposition: it should focus on the mechanisms that trigger social norms that involve, at the same time, given interests (that vary significantly themselves, and are linked to various representations and sometimes depend also on social norms) and social pressure within social groups or networks that make norms "social". We should find out why there are norms, and why they tend to be followed sometimes on the basis of and sometimes in opposition to given individual interests (which themselves, again, tend to vary often according to social norms).

I will propose in the first section a simple typology of the norms that define their different varieties and functions, insisting in particular on whether they tend to satisfy the interests of all the actors involved or, on the contrary, of some of them against others, since both situations are very common. I will then describe, in the second section, a first mechanism for triggering norms: the emergence of cooperation norms in social dilemma situations. These norms are based on symmetrical positions in social life and tend to lead to reciprocity norms. However, since there are asymmetrical positions in social life (where the actors do not have the same interests or the same power in the interaction) that can make room for cooperation, in the third section, I will explore the possibility of what I will name "asymmetric cooperation norms". I will describe then in the fourth section a second mechanism whereby cooperation social norms tend to lead to the setting of "group norms", based on shared cultural norms, that signal the belonging to specific cooperative groups. Since belonging to a given group increases the probability that people within those groups will respect cooperation norms, because belonging to a group increases the probability of sanctioning deviant behaviour, sharing the same cultural norms that signal such a belonging will tend to reinforce cooperation among members. Thus, different cooperative groups are set up on the basis of such group norms. This leads to the problem of the relationships between those different groups. I will describe then in the last section a third mechanism, whereby groups, being interpreted as coalitions, involve additional norms that impose to the members a submission to what is seen as group interests.

Pierre Demeulenaere - 9781789906851

1. BASIC TYPES OF NORMS AND THEIR RELATION TO INTERESTS

There are norms in most domains of social life; they correspond to the rules that tend to be followed in the various domains of social interaction that are "socially" supported or imposed within "groups". Someone can impose on oneself the following of a given rule, but in this case it is not a social norm as long as the person is alone in supporting the rule.

There are two basic types of norms. The first one corresponds to formal norms that are supported by specific institutions that have precisely the role of implementing the respect of the norms; the second corresponds to informal norms in the sense that there are no institutions devoted to enhance them, although the norms can be very accurate, precise and demanding. The first type corresponds therefore to "legal" norms or institutional norms, whereas the second type is more specifically linked to the sheer idea of "social norms" because they are diffuse and only reinforced by individual actors in the course of their interaction. However, from a sociological point of view, although legal norms and social norms are not at all similar in their functioning, their setting and the way they are reinforced, they are still both "normative" (in the sense that they prescribe certain types of attitudes and rules of behaviour) and "social" (since they are shared within communities). There is, therefore, no definite separation between legal norms and social norms which tend to interact (Posner, 2000). Thus, in the British constitutional system unwritten rules are very accurate and as widely respected as legal rules, although there are sometimes uncertainties regarding their exact scope and can be considered social norms depending on the willing of people to follow them. Legal norms and social norms sometimes overlap, they can be disjointed and they are sometimes opposed. This means that they fundamentally concern the same social normative issues despite the differences in the way they are managed. In addition to legal and social norms, there are also organisational rules which are set up within specific organisations. They are specific but, on a general basis, they should not be opposed to legal and social norms since they tend either to reflect them or are compatible with them. All three kinds of rules constitute the normative side of social life. They tell people what they should do in recurring typical circumstances.

It should be added that the notion of rationality itself has an obvious "normative" content, since it tells us the "right" thing we should do in specific choices (Weber, 1978), much beyond a simple cost-benefit analysis. Therefore, a debate has arisen: should rationality be considered as a cultural social norm among others or, conversely, as a normative governing instance that operates through the functioning of the human mind (Boudon, 2011; Mercier and Sperber, 2017), independently of variable cultural and social norms? This is a central debate that will not be developed here, although it is directly related to the analysis of the emergence and acceptance of social norms.

What is the function of social norms? They have two primary obvious functions that have been highlighted in the classical literature.

The first is a coordination role. The best example of this is the language, the way it has been initially described by Durkheim: people, in order to communicate, have to respect a common set of rules, otherwise they would not understand each other. In order to communicate, they *must* respect common standards; this is why a given language is a set of *social norms*. Another obvious example is the setting of standard weights and measures.

The second primary role of social norms derives from the first: coordination potentially leads to a reduction of conflict in the sense that, when norms are shared, in principle, they

Pierre Demeulenaere - 9781789906851

allocate roles and capacities of *legitimate* action and, therefore, reduce disputes about action decisions. Social roles defined by norms allocate resources for action, and thus determine who should do what in given circumstances.

However, an indirect effect of those two functions of social norms is to distinguish groups or individuals who do not share the same norms, who can be separated and sometimes opposed on this basis. Therefore, social norms can sometime lead to conflict among groups instead of coordination and reduction of conflict. I will come to this topic later in the chapter.

Should we contrast social norms with moral norms (Elster, 2009)? This is not necessarily a useful approach, because very often there is no clear separation between the moral and the non-moral aspects involved in the same norm; in particular, it is difficult to find specific emotions that are specialized in a separate moral aspect since there is an interplay between the various dimensions of norms, even from an emotional point of view. For instance, language has at first sight no obvious moral content, since its aim is only to coordinate actors when they speak or write to each other. However, language often involves, for instance, gender conceptions or racial prejudices, and when they are revealed or denounced language clearly becomes a moral issue and there are conflicts regarding the norms that should govern it. Moral issues intervene when individuals or groups feel they are hurt by a norm, and they will, therefore, sometimes try to change the norm. Since it is very common for people to feel hurt by given norms and, consequently, try to challenge them, it should be clear that social norms never stand far away from moral considerations (which themselves conversely rely on social norms, for instance, the idea that people should be treated equally or unequally). An essential feature of norms is indeed to assess who should be hurt or not hurt in specific ways, and in the "right" way. Property rights are, for instance, a way of determining whom others can "legitimately" hurt through economic competition (Alchian, 1987).

The possibility of coordination is an advantage, since people look for such coordination, which is necessary to everyday life. It would be impossible to communicate if everyone had their own private language. When people lack the ability to communicate with others, they cannot cooperate with them. Coordination, moreover, reduces partly the role of conflict, but this can be also done by hurting some other people if they find themselves submitted to the will of others through the norms (the way slaves are, for instance, or when a given language is imposed on others who would prefer not to speak it). Therefore, it is important to determine whom the norms benefit, and whom they harm.

Before addressing this issue, it is important to note that there might be strong divergences regarding the very notion of "harm" or "benefit". There are social norms that sometimes aim to precisely define what a harm is and what a benefit is: for instance, should people be allowed to smoke marijuana because some like it and consider it harmless, or should it be forbidden because it should be denounced as a harmful practice for oneself and for others? This is why neo-classical economics does not in principle specify any substantive idea of utility and relates it to "preferences" that are mostly psychological and social (Sen, 1977). Therefore, the notion of "harm" is not an obvious one: sometimes it is universal and is likely based on natural attitudes; for instance, in an epidemic, where it is assumed that there is a general desire not to be contaminated (although it can be set in conflict with the opposite desire of leading a normal life), but often it is related to the representations of specific groups; for instance, the desire to impose on women foot binding in classic China was considered to be a desirable thing (Appiah, 2010) on the basis that it expressed higher status. In this chapter, when I refer to "interests" I will not designate some kind of selfish utilitarian

Pierre Demeulenaere - 9781789906851

motives, but all the motivations that are pursued by actors in social life, which can appear to be complementary or opposed.

Regarding the question of whom norms favour or disfavour, there are fundamentally three possible situations commonly met in social life. The first is that everyone benefits from a given norm. There are many examples of that, and they are based on shared conceptions of advantage. Second, there are some norms that clearly favour some people and disfavour others. Again, there are many examples of that, and usually they also rest on shared conceptions of advantage, the extreme case being a norm in favour of slavery, where slaves are supposed to provide to their owners services they would prefer not to produce. Finally, there are norms that harm everyone involved in the following of the norm (Elster, 1989). However, this situation depends to a large extent on the two previous models because it is usual for people to believe that such norms are in fact beneficial ones—and from another point of view they are considered to be harmful. Thus the norm of consuming the brains of dead relatives at funerals that can transmit a deadly prion disease (Henrich, 2016: 144) is not seen through this unknown effect by the people who follow the norm. It can be either that there are harmful effects that are unseen, or that there is a controversy about what is genuinely good or bad (for instance in the case of circumcision). In addition, some norms are followed because everyone falsely believes that others would like to follow them; that is, some people, and maybe everyone, do not like a norm but believe that others support it because they consider it beneficial. This point has been analysed in a formal way (Centola et al., 2005; Willer et al., 2009), insisting on the structural effects of the norm diffusion. I will not further develop those situations in this chapter, and instead focus on those norms that serve either a collective aim, or the interests of some at the expense of others.

2. THE EMERGENCE OF SYMMETRICAL COOPERATION NORMS IN SOCIAL DILEMMAS SITUATIONS

I will start with situations where norms serve common advantages that are accepted as such by the participants. A trivial and ordinary device is to set up norms against possible actions that are unanimously considered to harm others. This is extremely common in social life and is based on the fact that there are harms that are unanimously considered to be so. There is, however, a need for norms because people can have reasons not to act in a way that is harmless, or because they are not aware of it being harmful: norms prevent people from having harmful attitudes, for instance the norm of wearing masks in order not to contaminate others with a disease. Norms can be set up intentionally as in the case of laws. They can also emerge in an unintentional matter; for instance, in hot climates people tend to eat spicy food. The mechanism responsible for this is quite simple: people who eat spicy food tend to live longer because spices protect the food against pathogens. If others imitate them, because there is also a norm of imitating older people (based on experience), progressively everybody will end up eating spicy food, without deliberately choosing to do so, and without knowing why spicy food is protective of their health and why there is a reason to follow the norm (Henrich, 2016).

A major mechanism whereby a norm is set up in order to satisfy people's interests can be located in the emergence of norms in social dilemma situations. There are recurring patterns in social life that have come to be known and named as "social dilemmas". They are characterized by two basic dimensions: First, these are situations where, when everyone follows his/her own

Pierre Demeulenaere - 9781789906851

interest, defined by the situation concerned, everyone finds him/herself at the end in a worse situation than if s/he had respected a cooperation norm limiting the immediate and exclusive pursuit of his/her interest. Second, a demand therefore arises for the establishment of such a norm of cooperation, which makes it possible to resolve this dilemma, this demand leading or not to its realization (Ullman-Margalit, 1977; Coleman, 1990).

Two simple examples illustrate such extremely common situations in social life. The first is that of a crossroads. If all motorists enter an intersection simultaneously, excessive traffic will tend to block them all. A norm of alternating passage could then solve the problem for the benefit of all. It is clear, however, that this norm is not easily and necessarily implemented: generally, when the traffic lights are defective, or when there is a very heavy traffic, the crossroads tend to be blocked because motorists do not respect the relevant norms (such as not to enter a crossroads if it is blocked). Failure to comply with the norms leads them to find themselves in a situation where, each blindly following his/her immediate interest (entering the crossroads), everyone is losing out, since the crossroads is blocked.

The second example is a classic one (Schelling, 1960: two rival superpowers each seeking to destroy the other by using nuclear weapons, but if one is attacked, it has time to respond immediately and to attack the other in return, so that both will be destroyed and therefore be losers in this process. Consequently, a symmetrical and reciprocal cooperative norm of non-aggression tends to be set up. It was respected, by way of illustration, in the historical example of the Cold War.

There are many social situations that can be characterized as social dilemmas, for example, trade relations between two countries, or greenhouse gas emissions, or protective behaviour in an epidemic etc. Sometimes, different types of norms can derive from given social dilemmas. Oström (1990) has thus highlighted that, as opposed to common wisdom among economists, the so-called "tragedy of the commons" does not necessarily lead to "individual" property rights, but to social norms regulating the use of a common resource. The reality of social dilemmas is obviously not a discovery made by game theory, since the common pattern of these situations belonging to social life was identified a long time ago in the classical literature, starting with Aristotle and developed in depth by a theorist like David Hume (Hollis, 1994). Game theory formalizes the situation, gives it its name, and associates it with a mathematical expression which limits the relevant conditions for characterizing such situations. In the basic game theory modelling of social dilemmas, such as the prisoner's dilemma, the different positions of the different players are thus strictly symmetrical and interchangeable, while in actual social life this is rarely the case. I will later introduce situations where people do not have the same interests, and the dramatically important situations where they do not have the same power.

A situation of social dilemma can, therefore, lead the people involved in it to the establishment of norms, which make it possible to improve the situation of current or future participants. However, this implementation is by no way self-evident. Repeated situations of social dilemmas can lead to the perception of the advantageous nature of the establishment of such a norm; it can, consequently, be effectively achieved, or not, and thereafter in a stable manner, or not. Indeed, it again appears that, if others respect the established norm, everyone may have an interest in not respecting it. But if all do the same, the advantage of the norm is lost. It is then the implementation of sanctions that makes it possible to reinforce and guarantee compliance with norms. This amounts to the problem of the modalities of the implementation (or not) of these sanctions, which represent a new derived social dilemma:

Pierre Demeulenaere - 9781789906851

people may have an interest in not contributing to the sanctions, hoping that others will assume the burden. This does not necessarily imply "selfish" actors; on the contrary, there are norms targeted at sanctioning selfish behaviour considered to be deviant, and many people will be ready to sanction selfish behaviours and will prefer adopting an "altruistic" behaviour. However, it is a possibility in such interactions that some people will behave in a selfish manner. And it is common that they will retaliate against a behaviour that is perceived as selfishly uncooperative.

In social life, there are two main categories of sanctions, corresponding to the two categories of norms previously mentioned. Legal norms, set up by specialized institutions (typically the state and its agencies), put formal sanctions in place. Social norms, for their part, are taken in charge directly by members of social life, in their immediate interactions, independently of these institutions, and are deployed through the threat of diffuse sanctions directly assumed by the actors. Contemporary life shows the continuous co-presence of these two types of sanctions (which are positive or negative). On the one hand there are the laws that govern social interaction; on the other hand, there are social norms that define what to do or not to do, in specific groups, and which allow for sanctions of "private" origin, which range from an eyebrow to ostracism, and sometimes to the use of unlawful violence. Positive sanctions are very popular, when for instance people encourage virtuous behaviours through applause. Sanctions are themselves governed by norms.

Social dilemmas therefore tend to lead to the establishment of norms, either legal or social, and to sanctions that reinforce them. In such a context, actors repeatedly face a choice between three typical basic attitudes: (1) a cooperative attitude of respect of "reciprocity" cooperation norms, (2) an attitude of cheating (free riding, given the existence of reciprocity norms), and (3) an attitude of unilateral domination consisting in imposing on others one's interest.

These three basic attitudes can be localized in a "primitive" (Boehm, 2012) recurrent situation, that of hunting in hunter-gatherer societies, where hunters can cooperate, cheat (try to escape the drudgery of hunting by trying to get food anyway) or adopt an attitude of "domination" requiring others to hunt for them. An evolutionary scenario (Bowles and Gintis, 2011), within the framework of 300,000 years of hunter-gatherer societies, may seek to establish a link between these repeated hunting scenes and the selection of emotions favourable to cooperative participation norms, and emotions unfavourable to cheating or unilateral domination without reciprocity. The survival of the species implies the participation of everyone in the hunt (at least those who are able to participate), therefore discouraging the attitudes (cheating or domination) that could dissuade individual effort. There is, therefore, for adaptive reasons, room for resentment against cheaters, and an aversion for situations where one finds oneself dominated. It is possible to understand on this basis "altruistic" behaviour (Tomasello, 2016), that is the respect of norms of reciprocity.

It should be noted, however, that the adaptive argument presupposes such an aversion toward domination and cheating (which is itself a form of domination). It does not explain it as such: indeed, were the people accepting of being dominated, or indifferent about that, they would continue to participate in the hunt despite being deprived of any personal benefit. The evolutionary scenario, therefore, presupposes, instead of explaining, a prior aversion to the situation in which one is dominated and has resentment toward cheating from their victims. Such resentment against domination seems to also be present in natural life (De Waal, 2013: 17). At any rate, such social dilemma situations and such symmetrical cooperation norms play a significant role in social life. They do not rest only on shared interests, but

Pierre Demeulenaere - 9781789906851

also on a refusal to be unilaterally dominated by one part who takes advantage of another. It should be added here that there are different types of reciprocal norms that can be more or less demanding, such as a requirement to help others, or to make gifts.

The purpose of sanctions is to strengthen compliance with norms of reciprocity. We can see that they are moreover generally linked to emotions that function themselves as "internal" sanctions and threats of sanctions. Social norms are typically supported by emotions that belong to the human species, which is why it is relevant to seek to analyse their emergence in evolutionary terms (Turner, 2007), independently of emotions, that can support norms and are maybe sometimes directly responsible for them, such as in the case of norms forbidding incest (Turner and Maryanski, 2005); there are specific legal and social norms, accompanied by sanctions, which are intended to effectively guarantee respect for cooperative attitudes, and which can orient toward different emotions. For instance, in some societies women can not have children outside marriage, and they will be typically shamed for that. If the norms forbidding children outside marriage disappear, they will no longer be shamed for that, hence there is a clear dependence of emotions on social norms.

3. COOPERATION NORMS IN ASYMMETRICAL SITUATIONS

It has been mentioned that the basic formalization of social dilemmas situations relies on two major (usually implicit) assumptions: the identity of interests, and the equality of power. What happens when those two major conditions are not met? I will suggest here that there are three typical outcomes, involving again the possibility of cooperation, domination and cheating:

- symmetrical cooperation norms, of another sort than in a social dilemma situation, because here the various interests are different, but there can be a common interest in letting everyone follow separately their own specific interests
- asymmetrical cooperation norms, where there is a balance between various interests that are at the same time, to various degrees, complementary and opposed
- a more straightforward domination of one part on another, when for instance the language of one group is imposed on another, against its will

First, when interests are not the same, the situation can lead to cooperative norms of reciprocity, if every participant is ready to accept other participants following their own different interests (which can be, again, dependant on "values" like religion), if they reciprocally accept their own. For instance, non-smokers can accept smokers smoking if they do not pollute them remaining in non-smoking areas. This amounts to a cooperative reciprocity norm, where no one wants to be submitted to others' interests in situations where it is possible to manage that each will follow their own aims without harming others. This can happen even if one part has the means to impose on others its own will. However, it is often very difficult to set up such symmetrical reciprocity norms, because of practical constraints: for instance, if in a country some people favour nuclear power, and others oppose it, it will not be possible to resolve the conflict just by saying that each party will follow its own way, everyone respecting the other's choice, on the basis of a reciprocal norm, a decision must be made, that will

Pierre Demeulenaere - 9781789906851

inevitably affect everyone, supporters or opponents. In such cases, another norm tends to be referred to, the one of majority decision (when consensus is not achieved). It relies on an underlying reciprocity norm, which is that, if no consensus can be reached, and if it is not possible to separate actions, then a majority decision will be taken that will affect everyone.

This leads us to situations where people have unequal power. Again, they have the choice between two major strategies: cooperation and domination.

Cooperation norms can arise when there is an advantage for it, based for instance on division of labour, that define different complementary roles, or the distinction between ordinary people and the people who are in charge of implementing sanctions. In such cases there tend to arise what can be named *asymmetrical* norms of reciprocity, which define different complementary roles and correspondingly different emotions supporting proper attitudes. For instance, in an organisation where there are unequal roles, where one gives instructions to someone who is supposed to follow them, there will be norms attached to obedience and deference (Blau, 1964) on the "inferior" side, and protection and moderation (not to abuse from one's power) on the "superior" side. However, it is often the case that when there is unequal power among individuals or groups, there is a tendency to sheer domination of one party over the other, because it is possible. In such situations there is no more cooperation as such, and there is a possibility of a complete absence of norms regulating the interaction, like in a pure state of war (or what Hobbes named a state of nature antecedent to any emergence of norms). But it is rarely the case. Even in situations of pure domination, like the relation between masters and slaves, there is a tendency to the emergence of asymmetrical cooperation norms, because what masters expect from the slaves cannot be obtained only by strength (slaves control some aspects of the situation). Therefore, in asymmetrical situations, there is still a tendency to the emergence of asymmetrical cooperation norms. The more people will share common interests, the easier it will be for them to follow those norms despite their unequal position. It is the case in corporate life, where shareholders, executives, employees, and stakeholders have interests that are partly complementary and partly opposed, and can lead either to symmetrical cooperation norms or to asymmetrical cooperation norms, or to sheer domination of one party by another.

In asymmetrical situations, there is a structural tendency to conflict and disagreement, because there is at the same time room for cooperation norms and for domination based on power asymmetry. In such settings, norms can contribute to stabilize relationships, notably through the idea of "proportional" contributions that seem to be a universal requirement (Boyer, 2018), although this notion of proportionality is often very hard to translate in accurate measures. This issue leads to three typical social situations; one puts a strong emphasis on hierarchical norms (in private life, in economic life, in political life, and so on) that attempt to transform unequal power positions into legitimate norms of interaction. It is the case of most traditional societies. Other types of societies are much more egalitarian and rely mainly on symmetrical egalitarian cooperative norms, which can be said to arise in social dilemmas situations. It is notably the case of most hunter-gatherer societies, exemplified by the Nuer as described by Evans Pritchard (1940), although some of them include hierarchical norms. Finally, contemporary societies offer a mix of egalitarian cooperation reciprocal norms in certain domains (like gender roles), and some room for unequal norms, mostly regarding income, wealth, and status, that are supposed to be attained through personal achievement (that can be referred to as reciprocity norms), but that are constantly challenged in the name of more egalitarian norms, on the basis of the largely documented inequality of opportunity.

Pierre Demeulenaere - 9781789906851

It is therefore important to assess the factors and mechanisms responsible for the evolution of historical norms regarding the extent to which important inequalities are accepted or not (Inglehart et al., 2005; North et al., 2009).

It is also important to understand that people can have a sense of justice that will prevent them from unilaterally dominating others, even if they are able to do so (Moore, 1978). People tend to resent unilateral domination exerted against them, and therefore can understand that others are in the same position and will not be ready to accept unilateral domination. This is a point that had already been somehow made by Adam Smith (1976). However, as noted by Weber, people who are in a dominant position will generally tend to believe that it is a legitimate one and that their interests are legitimate. It is the case when right-handed people impose on left-handed ones an obligation to change their way of writing. Here it is probably an idea of "normality" that prevails, as opposed to the reciprocal situation previously mentioned, where both right-handed people and left-handed people would be allowed to behave the way they tend naturally to do.

4. THE ROLE OF CULTURAL NORMS IN REINFORCING COOPERATIVE SOCIAL NORMS

Norms of cooperation (symmetrical or asymmetrical) are a central feature of social relationships. No society can avoid having such norms, although there are historical differences regarding the importance of their respective role. There are historical variations in the way those norms are implemented, depending on various factors and mechanisms.

In this regard, an essential point is to emphasize that the cooperative space that norms allow is in principle unlimited, potentially including all direct or indirect participants in an interaction. In other words, the interest in cooperation and the need for sanctions, in order to reinforce them, do not specify and do not require any direct separation between one group of partners of the interactions and others; it is potentially "universal", integrating all the actors who could possibly be involved in such interactions. The concept of cooperation in situations of social dilemmas does not directly imply separation into groups and the relevant social norms are fundamentally extensive.

However, things do not happen that way, for essentially practical reasons. In practice, the exercise of sanctions reinforcing norms is more specifically directed at a *limited* group of people who are likely to be targeted by the potential sanctions put in place. The imposition of sanctions takes place within the limits of an established set of people and thus constitutes, to a large extent, the group in question, which is then characterized as a guaranteed cooperative space. There are two sources, therefore, of the targeting of norms and sanctions. Institutions, typically states, formally define (more or less precisely depending on the degree of rationalization and formalization) a population and its limits, characterizing the persons affected by the norms and their associated sanctions, and thus constituting them as a group concerned by the application of these standards. This closure, depending on historical conditions, is largely arbitrary. Social norms do the same: they tend to circumscribe groups to which the norms are applied through possible sanctions. However, as opposed to institutions, which, once they are set up, have the means of defining the limits of a group defined by them, social norms do not directly have the same ability to do so. Since cooperation social norms are extensive, they need a specific means to be reinforced.

Pierre Demeulenaere - 9781789906851

The idea I will now defend, in this perspective, is that, in addition to *extensive* cooperation social norms of reciprocity, there are *particular* cultural norms of coordination, obeying their own logic, that can make it also possible to *signal* (Gambetta, 2011), by their particularity, the belonging to a specific group collectively and effectively governed by norms and thus guaranteeing the higher probability of a cooperative attitude. The sharing of common cultural norms makes it possible to signal belonging to a group in effective interaction where the norms are more likely to be respected, because there is then a higher probability of sanction than toward "strangers" to this group. Thus particular cultural norms (a particular language, a particular religion, food habits, clothes, table manners, etc.) can achieve, in addition to their own purposes, which are coordination functions in the social interaction, the result of ensuring the guarantee of a cooperative group and therefore "trust" among their members.

These norms are generally not directly intended to constitute a particular group distinct from other groups and separate from them. They aim at coordinating practices and beliefs and can therefore have themselves an indefinite extensive character, the coordination of action being able to relate to any interaction: a language, as a set of social norms, is not intended to remain necessarily confined to a particular group, it can possibly be extended to all potential participants in the interactions.

However, since there are several possible given norms of coordination available and that can be identified historically, and which evolve separately on their own basis, they can be identified and associated, at given times, with particular cultural groups. They become "cultural" norms. These are then rigidified and limited because they are superimposed on cooperative groups which are themselves, for the reasons we have seen (definition of a group where norms and sanctions are exercised), limited. Particular physical traits can also be and frequently are associated with particular groups; they too can then serve as a "signal" of belonging to a cooperative group. These physical traits are then integrated by social norms into group traits and thus become paradoxically cultural traits.

The typical mechanism for the emergence of group norms can be described as follows:

(1) Situations of social dilemma raise the need for extensive norms of cooperation.
(2) The norms of cooperation and reciprocity (symmetrical or asymmetrical) tend to be established, but they are fragile in the face of the risks of cheating and of unilateral domination.
(3) Threats of sanction increase the probability that the reciprocal norms of cooperation will be respected.
(4) These threats of sanction are, for practical reasons, carried out on a target population, circumscribed as a cooperative group, where the chance of cooperative behaviour is reinforced by the threat of the exercise of sanctions related to social norms.
(5) Consequently, a distinction is made between a given cooperative group and its external environment: one has the possibility of having confidence in the members of a cooperative group to which they belong; outside of the group such a guarantee does not exist. Group limits are thus established, either by formal norms linked to the establishment of institutions specialized in the management of sanctions, or by diffuse social norms aimed at strengthening cooperation.
(6) On this basis, particular cultural norms of coordination can serve additionally as a "signal" of belonging to the cooperative group in question: cultural particularities indeed manifest the singularity of a distinct group, whose members have, by sharing social

Pierre Demeulenaere - 9781789906851

norms, a greater chance of adopting cooperative behaviour toward each other, because deviance is more likely to be sanctioned among them.

(7) The particularity of cultural norms is in itself independent of the closure achieved, on a target population, by the threat of sanctions making it possible to strengthen cooperation.

(8) It happens that the closure operated by the space of social norms on a target population is symbolized by the singularity of particular cultural norms (or particular physical traits), which function as a signal of belonging to the cooperative group.

(9) There is therefore an interaction between the norms of cooperation and the cultural norms of coordination, which will be reinforced in their respective exercises.

(10) Social norms, finally, by the closure (even temporary, pending future reconfigurations) that they operate, tend to create norms of collective belonging to the group, to institute a particular culture (Gellner, 1983) and therefore to arouse feelings of collective belonging (Anderson, 1983). Cultural norms are therefore defined by their exclusivity.

There is a large body of literature on the emergence of social norms in situations of social dilemma. There is also considerable literature on the fact that relationships of trust are strengthened within a particular cultural group (Granovetter, 2017). The mechanism described distinguishes the two dimensions, and makes the link between them: it shows how, while the logic of cooperation is in itself potentially indefinitely extensive, and while cultural norms (obeying their own logic of coordination) do not in themselves have a direct vocation to constitute closed groups, since their logic of coordination can as well be extensive, the threat of cheating or domination in situations of social dilemma leads to the establishment of cooperation social norms, which require the possibility of sanctions; these are exerted on a target population. The specificities of cultural norms, when they are present, make it possible to effectively signal the existence of such a cooperative group and reinforce cooperation norms. Cultural norms can, moreover, be deliberately instituted in order to create a cooperative space based on homogeneous coordination standards.

The size of these groups is fundamentally variable: it can range from a population of over a billion inhabitants unified by common institutions, common social norms, common cultural norms, and common ethnic traits to micro populations of a few hundred people united on the same bases. But there is always a tendency to set up subgroups within groups formed from a certain point of view. This can be found in hunter-gatherer societies (Apicella et al., 2012). Conversely, given groups tend to cooperate with each other and to set up norms of cooperation beyond their internal limits.

On the basis of the conjunction between social norms of cooperation, groups of people subjected to the threat of sanctions, and specific cultural norms develop feelings of collective belonging generally marked by "solidarity." The emotion of solidarity corresponds to this feeling of unity in a group where trust relationships prevail, because the effectiveness of social norms guarantees it. In addition, there is a tendency to develop in such circumstances the *belief* in the existence of a "group" heterogeneous with its members (Weber, 1978), since the traits (that is the norms) that constitute the group tend to survive to their members. This belief is relevant, since the groups are empirically identifiable and are distinct from one another, because of a particular conjunction of cooperative norms and of cultural norms; and because the norms can be identified independently of the individuals who share them. On the other hand, it is wrong to believe that there is a "naturalness" of these groups, or a perennial historical essence, or a

Pierre Demeulenaere - 9781789906851

stable cultural identity because they are always the result of variable normative configurations, fundamentally evolutionary and in this sense precarious.

5. GROUP PRESSURE AS COALITION NORMS

It is very common, in social life, to find norms that oppose groups. The mechanism previously described does not explain that; it only explains norms of belonging to groups, not the norms related to their opposition. I will describe now a mechanism that triggers such norms related to the opposition of groups, which are then understood as coalitions exerting pressure on their members in order to achieve what are seen as groups' interests against other groups.

A cooperative group subject to institutional, social, and cultural norms is not, as such, a "coalition" opposed to other coalitions: the difference is conceptual and practical—and emotional. A solidarity group is not directly a coalition insofar as a coalition implies, in addition to the cooperation norms, and their implementation in definite groups, the existence of a common goal, which must be achieved in competition with other groups' goals, or other individuals' goals, possibly identified as common adversaries. These goals, and the competition between them, often do exist, and the group becomes then a coalition against these adversaries. It leads to the implementation of supplementary "coalition norms" that are not only symmetrical or asymmetrical reciprocity norms, but introduce devotion to the "group" and pressure on individuals who compose it, when it is opposed to other groups. This can be seen for instance in the competition between companies that try to reinforce a sense of belonging through corporate cultures (Akerlof and Kranton, 2010).

It is not the existence of groups as such which is the source of conflict among them (because they can also enter a cooperative relationship), but the opposition of interests, which, already present in the interaction of individuals (concurrently with the very frequent complementarity of their interests that lead to cooperation norms), can also intervene in relationships between organized groups once they are settled by norms and are considered as quasi-actors (Coleman, 1990). Opposition of interests frequently arises between individual people. They are, to a large extent, unsurpassable, because of the existence of "rival goods", which are such that the advantage of one can be gained at the expense of the other. These "goods" are not necessarily "material"; they can be cultural, symbolic, or ideological. In the relatively simple framework of a social dilemma, and the emergence of norms of cooperation, the potential opposition of certain specific interests (those linked to an attitude of cheating when being a free rider) tends to be limited if the cooperation institutional and social norms are effective. However, all conflicts of interest, linked to the asymmetry of positions, to the existence of rival goods (and the complexity of these), and to the divergence of interests and representations, are far from being settled by such devices. Their surge is permanent, and there is no reason to believe that we can reach a situation where they would once and for all disappear.

Thus, since groups tend to be seen as individuals themselves, they can enter conflict relationships to the extent that individuals do. But since they are made of several individuals, groups norms will intervene that will exert a pressure toward realizing a stronger cohesion among them.

What then is the relationship between one group and other "external" groups? There are basically three possibilities, which are the same as those existing in social interactions. The first is the extension of cooperation, corresponding to the establishment of reciprocity

Pierre Demeulenaere - 9781789906851

norms (symmetrical or asymmetrical). The second is cheating when cooperation norms have been previously adopted or are implicitly taken for granted in a partnership. The third is the attempt at unilateral domination, regardless of any cooperation norms (although they can also lead, more or less inevitably, to asymmetric cooperation norms). Relationships between groups present these three possibilities and are therefore analogous to those that exist in basic interaction between individuals.

Coalitions then allow the exercise of obtaining an advantage for a group at the expense of another one: either, overall, to achieve domination, or cheating in the interaction with another group; either to protect the members of a given group against threats of domination or cheating emanating from another group. In reality, these two orientations tend to mix the opposition of interests causing mistrust, conflict, and then the "justification" in the eyes of aggressors of their aggression by reference to the threat of an aggression from the other group.

This can happen between different groups established on the basis of institutional norms, but also between groups unified by sharing common social and cultural norms, independently of such institutional devices, like social classes in the sense of Weber. This can also concern groups defined by only particular economic interests (fishermen, for example) who enter an alliance to defend their interests and thus constitute themselves as a pressure group. Conflicts therefore arise over conflict of interests between groups. These conflicts can sometimes be replaced by extensive cooperation (symmetrical or asymmetrical). But, once groups are formed, and identified with individuals, these conflicts tend to arise because there are rival goods, in the same way that there is a tendency to conflicts between individuals, partly resolved by cooperation norms, but largely unresolved. It stems always from the same situation: the advantage of one group can be achieved at the expense of that of another, when one is in a zero-sum game or in the competition for the obtaining a rival good.

Historically, these conflicts between groups have repeatedly concerned typical zero-sum games: the annexation of land that is a source of wealth; the conflict over control of resources that can be unequally shared; the conquest of external populations, subject to tribute or enslavement; the search for the pre-eminence of particular cultural norms, if these are perceived as incompatible with others and superior to them. When cultural norms are perceived as incompatible, they correspond in a way to a rival good, where the success of one corresponds to the defeat of the other. Today, the conflict between groups has renewed sources: for example, the definition of trade rules between two nations that are hardly perfectly cooperative and symmetrical, or the number of migrants allowed entering another country (on the basis of the distinction of groups). Within a group, subgroups can also request changes in their favour: once again, either to obtain an advantage at the expense of others, or on the contrary to protect themselves against other groups which seek to get benefits at their expense.

These conflicts between groups therefore correspond to the reality of a zero-sum game in which obtaining a rival good presents the possibility of winners and losers. This reality may be mistakenly perceived, and one may believe in the existence of a zero-sum game when it would be in fact possible to be in a positive-sum game. There are simple cases where, indeed, all the members of a group A can be in conflict with all the members of a group B. Thus, if all the members of a group A seek to reduce slavery in all the members of a group B, and if the latter do not wish to be enslaved, we can consider that there is both a conflict between group A and group B, and a conflict between all members a of group A and all members b of group B, the different participants in the interaction being clearly identified as belonging to one or the other of the two camps.

Pierre Demeulenaere - 9781789906851

We can see that in such circumstances there is a second type of "solidarity" that intervenes, additional to the basic solidarity emanating from a cooperative group based on trust: the solidarity resulting from the constitution of an alliance against another group. This solidarity typically strengthens the feelings of belonging to the group. It leads hereafter to four other norms (and corresponding emotions):

- Hostility toward members of other groups, which are viewed as interchangeable foes
- Hostility toward the members of one's own group when they are suspected of being traitors to the coalition
- Encouragement to sacrifice toward the group seen as "superior" to each of its members, on the basis of the belief of the existence of such a group
- Importance of cooperation between the members of the coalition, regardless of their hierarchy or potential internal conflicts (Boyer, 1993, 2018).

Those group coalition norms aim at homogenising the members of each group, by defining them as symmetrical partners, despite their differences, and equally submitted to the same group "interest". It has been empirically observed by Weber that such group membership tends to lead to a typical opposition between internal ethic and external ethic. Anthropologists (LeVine and Campbell, 1972) signal the importance of a tendency toward ethnocentrism; it is based on the inner mutual reinforcement of institutional, social, cultural and coalition norms that strengthen the separation between groups.

However, such inner homogeneity and such homothetic relationships between members of rival groups rarely exist unequivocally because of the complexity of interest relationships and because of the proper logic of all the norms involved, that can also lead to cooperation. For example, in trade relations between two countries, or in determining the legislation of a given country, and if only financial interest is taken into account, it is extremely difficult to find rules which will simultaneously satisfy all sectors of activity and all participants in them. Regulatory decisions often tend to seek rules that suit the "majority" of participants, or conversely rules are sometimes put in place in favour of minorities against the majority. Consequently, the "losers" of these rules will appear themselves as a particular group (fishermen, for example) and will tend to have their cooperative solidarity reinforced by alliance norms against the beneficiaries and promoters of the rules. If the subgroup in question (fishermen, for example) also corresponds to a subgroup unified by particular cultural norms, the feeling of identity and persecution in relation to a dominant group will be reinforced. Solidarities then occur at several normative integrative levels, and emotions can manifest themselves concurrently depending on the levels concerned.

The exercise of such coalitions is linked to emotions that allow norm reinforcing, as previously. These emotions are part of the behavioural equipment of human beings. It is possible that they are selected as a result of a natural evolution process promoting cooperation, since we do not choose to have emotions. However, the question here is whether the reality of group-belonging norms stems directly from specific natural emotions, orienting us toward group belonging, whatever the group, as argued for example by Maynard Smith and Szathmary (1999), Greene (2013), and Haidt (2012), following earlier works by Sherif (1966) describing the psychology of group belonging. The idea would be that we tend by nature to belong to groups, on an unconscious basis, developed through natural selection. If, analytically group norms can be distinguished, corresponding to two types of "solidarity", cooperative solidarity

Pierre Demeulenaere - 9781789906851

(which does not involve rivalry with another group, but that can be reinforced through sharing cultural norms), and coalition solidarity (intervening against another groups and implying the idea of a submission to the group's interests), the associated emotions are nevertheless basically the same and correspond to the basic emotions of social interaction. There are "positive" emotions oriented toward cooperation (symmetrical or asymmetrical), when it is in one's favour, and "negative" emotions oriented toward cheaters, internal or external to a group, or toward suffered domination, internal or external to a group. I will not however discuss here the proper evolutionary reasoning about the selection of group norms.

However, since group norms and the definition of group limits tend to evolve, on the basis of extended possibilities of cooperation, it does not seem relevant to give those norms a stable natural basis. In this chapter, I have outlined a mechanism for the emergence of group norms that does not need any kind of specific natural emotions to support the explanation of such group feelings. On the basis of the knowledge of such a mechanism, a rational approach to norms justification can lead to a plea in favour of extension of cooperation, against a narrow defence of group belonging and closure. It is the case: very often people prefer universalistic norms against narrower norms limited to their own group. This is because the possibility of cooperation is fundamentally extensive, and groups are set up in order to reinforce cooperation.

CONCLUSION

In this chapter, I have tried to analyse three major mechanisms triggering norms, and the different types of norms that they produce. I have insisted on cooperation norms that are triggered in social dilemma situations, and that are sometimes symmetrical and sometimes asymmetrical. I have then described a second mechanism that leads to the closure of groups in order to reinforce norm following. Various cultural norms allow then signalling the reinforcement of the probability that people, within the groups, will follow cooperation norms. And then, when different groups are set up, the issue of their relationships tends to transform them into coalitions that impose supplementary norms on their members. At each step, there is a complex interplay between interests and norms, the general device being that, in order to ensure that given interests are satisfied within interaction between social actors, norms are set up that tend to limit the possible choices for individual actors. Therefore, norms, at the same time, derive from interests and limit and define them in a drastic way.

REFERENCES

Akerlof, George A. and Rachel E. Kranton (2010), *Identity Economics: How Our Identities Shape Our Work, Wages and Well-Being*, Princeton and Oxford: Princeton University Press.

Anderson, Benedict (1983), *Imagined Communities: Reflections on the Origin and Spread of Nationalism*, London and New York: Verso.

Alchian, Armen A. (1987), "Property Rights", in the *New Palgrave Dictionary of Economics*, New York: Norton.

Apicella Coren L, Frank W. Marlowe, James H. Fowler and Nicholas A. Christakis (2012), "Social networks and cooperation in hunter-gatherers", *Nature*, **481** (7382).

Appiah, Kwame Anthony (2010), *The Honor Codes*: *How Moral Revolutions Happen*, New York: Norton.

Blau, Peter (1964), *Exchange and Power in Social Life*, New York: Wiley.

Boehm, Christopher (2012), *Moral Origins*: *The Evolution of Virtue, Altruism, and Shame*, New York: Basic Books.

Pierre Demeulenaere - 9781789906851

Boudon, Raymond (2011), "Ordinary Rationality: The Core of Analytical Sociology", in P. Demeulenaere, *Analytical Sociology and Social Mechanisms*, Cambridge: Cambridge University Press, 33–49.
Bowles, Samuel and Herbert Gintis (2011), *A Cooperative Species. Human Reciprocity and its Evolution*, Princeton, NJ: Princeton University Press.
Boyer, Pascal (1993), *The Naturalness of Religious Ideas*, Berkeley, CA: University of California Press.
Boyer, Pascal (2018), *Minds Make Societies. How Cognition Explains the World Humans Create*, New Haven, CT: Yale University Press.
Centola, Damon, Robb Willer and Michael Macy (2005), "The emperor's dilemma: A computational model of self-enforcing norms", *AJS*, **110** (4): 1009–40.
Coleman, James S. (1990), *Foundations of Social Theory*, Cambridge, MA: Belknap Press.
De Waal, Frans (2013), *The Bonobo and the Atheist*: *In Search of Humanism Among the Primates*, New York: Norton.
Elster, Jon (1989), *The Cement of Society. A Study of Social Order*, Cambridge: Cambridge University Press.
Elster, Jon (2009), "Norms", in Peter Hedström and Peter Bearman (eds.) *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press, 195–217.
Evans Pritchard, E.E. (1940), *The Nuer*: *A Description of the Modes of livelihood and Political Institutions of a Nilotic People*, Oxford, UK: Clarendon Press.
Gambetta, Diego (2011), *Codes of the Underworld: How Criminals Communicate*, Princeton, NJ: Princeton University Press.
Gellner, Ernest (1983), *Nations and Nationalism*, Oxford, UK: Basil Blackwell.
Granovetter, Mark (2017), *Society and Economy*: *Framework and Principles*, Cambridge, MA: Belknap Press.
Greene, Joshua (2013), *Moral Tribes*: *Emotion, Reason, and the Gap between Us and Them*, New York: Penguin.
Haidt, Jonathan (2012), *The Righteous Mind*: *Why Good People Are Divided by Politics and Religion*, New York: Vintage Books.
Hechter, Michael and Karl-Dieter Opp (eds) (2005), *Social Norms*, New York: Russell Sage Foundation.
Henrich, Joseph (2016), *The Secret of Our Success*: *How Culture is Driving Human Evolution, Domesticating Our Species and Making Us Smarter*, Princeton, NJ: Princeton University Press.
Hollis, Martin (1994), *The Philosophy of Social Science*: *An Introduction*, Cambridge: Cambridge University Press.
Inglehart, Ronald Christian Welzel (2005), *Modernization, Cultural Change and Democracy*: *The Human Development Sequence*, Cambridge: Cambridge University Press.
LeVine, Robert A. and Donald T. Campbell (1972), *Ethnocentrism*: *Theories of Conflict, Ethnic Attitudes, and Group Behavior*, New York: Wiley.
Maynard Smith, John and Eörs Szathmary (1999), *The Origins of Life*: *From the Birth of Life to the Origin of Language*, Oxford, UK: Oxford University Press.
Mercier, Hugo and Dan Sperber (2017), *The Enigma of Reason*: *A New Theory of Human Understanding*, Allen Lane, Penguin.
Moore Jr, Barrington (1978), *Injustice*: *The Social Bases of Obedience and Revolt*, London: Macmillan.
North Douglass C., John Joseph Wallis and Barry R. Weingast (2009), *Violence and Social Orders*, Cambridge: Cambridge University Press.
Oström, Elinor (1990), *Governing the Commons*: *The Evolution of Institutions for Collective Action*, New York: Cambridge University Press.
Posner, Eric A. (2000), *Law and Social Norms*, Cambridge, MA: Harvard University Press.
Schelling, Thomas C. (1960), *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.
Sen, Amartya (1977), "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory", *Philosophy and Public Affairs*, **6** (4): 317–44.
Sherif, Muzafer (1966), *The Psychology of Social Norms*, New York: Harper Press.
Smith, Adam (1976), *The Theory of Moral Sentiments*, Oxford, UK: Clarendon Press.
Tomasello, Michael (2016), *A Natural History of Human Morality*, Cambridge, MA: Harvard University Press.
Turner, Jonathan H. (2007), *Human Emotions*: *A Sociological Theory*, London and New York: Routledge.
Turner, Jonathan H. and Alexandra Maryanski (2005), *Incest*: *Origins of the Taboo*, London: Paradigm Publishers.
Ullmann-Margalit, Edna (1977), *The Emergence of Norms*, Oxford, UK: Clarendon Press.
Weber, Max (1978), *Economy and Society*, Berkeley, CA: University of California Press.
Willer, Robb, Ko Kuwabara and Michael W. Macy (2009), "The false enforcement of unpopular norms", *AJS*, **115** (2): 451–90.

Pierre Demeulenaere - 9781789906851

13. Coupled contexts[1]

Filippo Barbera and Nicola Negri

INTRODUCTION

Contexts are the bread and butter of sociological thinking and sociologists often invoke them as the key factor at work. Despite its ubiquity, the analytical added value of the concept is far from clear (Boudon 2014). Sociologists rely on contexts for the explanation, description and interpretation of social phenomena; to formulate the right question and to give the correct answer; to draw the boundaries of the explanandum and to set the scope of the explanans. As Raymond Boudon put it: "the question as to 'What is context?' has actually no general answer, but answers specifically adapted to the challenging macroscopic puzzles the sociologist wants do disentangle" (2014, 43). From the analytical sociology viewpoint, contexts matter for the macro-micro-macro explanation of social phenomena, where, at the first step, the actions of individuals are explained and, in a second step, the collective outcome of the interdependent actions of many individuals are explained (Udehn 2001). In this line, analytical sociology has been following the J.S. Coleman (1990) "warning" to keep the micro-level model *as simple as possible*. While this warning did not imply per se the adoption of rational choice theory (RCT from now on), the assumed "contracted" definition of intentional behavior in terms of RCT has been deemed particularly fit for purpose by many social analysts who followed Coleman's advice (Raub et al. 2011). Contrariwise, in line with the analytical sociology mainstream (Hedström 2005), we will be arguing that the *contracted reductionism* of intentional behaviour to RCT might easily contradict the principles of analytical realism to which analytical sociology explicitly adheres. Nonetheless, our standpoint will crucially differ from the analytical sociology mainstream in one key aspect: we will hypothesize that even in "middle-level" complexity contexts the features of social structure need to be taken into account in *close connection* with the micro-level model.

We will hence assume that contexts, structures and micro-level logics of action interact and change *together* (Granovetter 2002). This implies a shift from natural to social rationality (Lindenberg 2003) and, at the same time, it requires dealing with the dwindle of analytical power that this shift implies. To this end, we will be arguing that the micro-level reference point of analytical sociology—the so-called Desires-Beliefs-Opportunities model (DBO)—still serves the purpose, providing that it is *not* employed as a propositional theory of action. Rather, it should be understood as a conceptual device standing in between a general framework and a propositional theory, which is, at least in principle, well-matched with a number of micro-level principles.

[1] To the memory of Nicola, my first mentor, an esteemed colleague and a dear friend.

Filippo Barbera and Nicola Negri - 9781789906851

Our argument is as follows. In the first section we will illustrate the lack of realism that connotes RCT in the analysis of "middle-level" complexity contexts as those regulated by vacancy-chains models and organizational demography. We will, additionally, illustrate the payoff that the DBO framework entails in terms of realism, while shifting from natural to social rationality. In the second section, we will furthermore suggest a number of ways of combining positions and social networks in organized settings to analyse the co-variations of interaction contexts, the social structures and individual action patterns. Finally, in the last section we will succinctly suggest that such a gain in analytical realism is key to explain "radical innovations" that *never* existed before.

1. NOT EVERYTHING WENT LIKE IN "MODERN TIMES"

To grasp the complexity of social structure a simple theory of action is required. This was Coleman's proposal in *Foundations of Social Theory* (1990). This recommendation has been successfully followed by many scholars, chiefly by those embracing the RCT approach (e.g. Hechter and Kanazawa 1997). From the standpoint of analytical sociology, we maintain, this proposal puts the claims of parsimony *before* those of realism and thus pushes toward the building of models equipped with great analytical power (few assumptions, many consequences), but not as effective in making sense of social phenomena in "middle-level" complexity contexts.

In this regard, let us consider the vacancy-chains used by Sørensen (1977) to study career dynamics in occupational structures. In the apt summary offered by Petersen (2004, p. 30):[2]

(1) There is a fixed occupational structure characterized by a distribution of jobs;
(2) Opportunities to get ahead will be determined by the rate at which vacancies in the occupational structure open up, through turnover, jobs creation, and the like;
(3) People differ in the job-relevant resources they possess (e.g. human capital, experience), which help them compete for positions in the occupational structure;
(4) People want to get ahead in the structure (the action part);
(5) The rate at which opportunities to get ahead in the occupational structure open up varies systematically between demographic groups. These variations depend on ethnic discrimination mechanisms, but others mechanisms—such as those based on gender discrimination—could be at work as well.

Points 1, 2 and 5 specify structural properties, while points 3–4 define individual-level properties. Accordingly: "From these five features characterizing a stratification system, (Sørensen's) analysis focuses on the immediate implication that individual success depends not only on (3)

[2] While using the Sørensen model, we must underscore that the issue at stake does not concern the functioning of careers. In particular, there is no attempt to deepen the analysis of how and why individuals have come to occupy certain positions, or have succeeded or not in the competition for vacancies. The reference to this model is used exclusively to spell out key hypotheses on the intertwining among the complexity of the social structures, the contexts of interaction and the micro-level model.

Filippo Barbera and Nicola Negri - 9781789906851

individual resources, but also on the (1) distribution of positions, (2) the rate at which positions open up, and (5) which group one is in and when one entered the labor market" (Petersen 2004, p. 30). This approach predicts the same concave career lines with respect to the time axis as human capital theory, but starting from opposite assumptions of a structural kind (Hedström 2004).

In accordance with analytical sociology principles, let us assume that structural properties and their dynamic per se do not *generate*, transform, produce or change anything. The role played by structural properties does not mean assigning it a generative power that is autonomous and independent from the micro-level generative causality (Hedström 2005).[3] The number of positions that will be filled at time $T + 1$ is clearly the effect (intended and unintended) of individual actions and institutional constraints at time T (Boudon 1973). Would it be feasible, on this ground, to assume that RCT is a proper tool to make sense of career choices in vacancy chain competitions?

To answer to this question, it may be helpful to briefly recall a vignette from the historical experience of the fordist factory in Italy in the 50s. An historical figure of Italian trade unionism, Vittorio Foa, recalled in his autobiography (1991, p. 280, our translation): "[in] the fifties we were very impressed by the witness of a French worker from Renault who illustrated that even in the work at assembly line, which seemed the most 'commanded' by the machines times and methods, nothing would march without an autonomous workers' initiative". Perhaps, from this viewpoint, the factory portrayed by Charlie Chaplin in *Modern Times* did not tell everything. As Burawoy's (1979) ethnographic experience in a piece-rate machine shop showed, management controls workers by giving them the "illusion of choice" in a highly organized environment. In the machine shop, the piece-rate system created the illusion of labor *as a game*. Workers competed with each other to "make out" and surpass their expected production quotas and this, over time, was a source for their job satisfaction. Accordingly, those more skillful gained the most *respect* and *prestige* from their fellow coworkers: "Until I was able to strut around the floor like an experienced operator, as if I had all the time in the world and could still make out [produce the quota], few but the greenest would condescend to engage me in conversation" (Burawoy 1979, p. 64, in Granovetter 2005, pp. 41–2).

In this vein, it is crystal clear that the involvement of workers and their active participation call for more complex micro-level motives than those of the standard incentive models based on the RCT, such as those that leverage on efficiency wages (Milgrom and Roberts 1992). Even within the fordist assembly line—explicitly designed to maximize the independence of performances from the subjective characteristics of workers—empirical behaviors cannot be reduced to the RCT axioms. A realistic explanation requires to include

[3] Even structurally oriented scholars do not deny the generative power of social action. Peter Blau, arguing for structural parameters, is forced to admit: "To state that parameters create differentiation is speaking elliptically, of course. Parameters are concepts for observing the lines of differentiation among people created in their *social interaction*" (1974, p. 617 emphasis added). And as Sørensen also pointed out, "[T]he appropriate causal language to be used in empirical work should be one of causal narrative, accounts of how the causal linkages are created … [A] model focusing either on change over time or change in utility, is indeed the proper basis for empirical work" (Sørensen, 1999, p. 10).

Filippo Barbera and Nicola Negri - 9781789906851

feelings, social identities and expressive behaviours of workers.[4] And this—to restate the key point of this paper—is not for the sake of micro-level realism per se, but to grasp the complexity of social structure.

If RCT is not a proper tool in a highly standardized environment as the fordist assembly line, it is even less so in the vacancy chain modeled by Sørensen. In fact, in the Sørensen model employees have to take into account that the odds of "winning" a position are influenced by a large number of dimensions: length of queues, organizational structure, people's age, internal and external labor market dynamics, "glass ceiling" effects and so forth. A much higher number of dimensions than those, for example, covered by an RCT incentives scheme. The cognitive burden required to process these dimensions according to a RCT model is quite demanding and, for sure, at odds with the requirements of analytical realism. It would be unrealistic—though technically feasible—to include all these dimensions as arguments in an expected utility function which computes the optimal career line in terms of preferences and constraints with regard to the lifetime career. Furthermore, if we would include all these dimensions as variables in a linear regression model, we would end up implicitly assuming a "gas station theory" of vacancy competition (Sørensen 1998). Namely, we would need to assume that employees compute their chances of winning on the basis of a great variety of information inputs, gotten from a very wide range of *independent* and *additive* dimensions. And this would be equally at odds with the requirement of analytical realism which entails that the process through which the outcome is generated should be represented in a realistic way (Sørensen 1998).

The weakness of rational choice accounts is even more explicit in the so-called "organizational demography" models, which have extended the idea of vacancy chain to include a variety of organizational structures (e.g. Stewman and Konda 1983; Stewman 1986). Here job distributions, vacancy chains, and managerial staffing and hiring practices generate a compound structure of "probabilities for advancement". These are the key elements to model the so-called organizational "Venturi tubes", which specify the relative career chances over the full organizational hierarchy. The "Venturi tubes" illustrate that individual promotion chances may take different forms and these are *not* isomorphic to formal organizational structure: "Pyramidal organizations often do not have pyramidal chances for promotion" (Stewman 1986). Therefore, individual promotion chances cannot be correctly inferred from formal organizational structure, and this prevents *even more* careers from being governed by RCT models.

At first glance, even Sørensen's analysis does not require micro-level assumptions that radically deviate from the RCT ones; points 3 and 4 of the Sorensen model mentioned above are coherent with this theory. This leads to a quasi-paradox: on the one hand, Coleman's suggestion is in many cases very effective. The "as if" analysis carried out in the wake of RCT often helps to make sense of the complexity of social structures (e.g. Raub and Weesie 1990). In this line, both economists and sociologists (Burt 1992) follow a RCT approach to understand how information flows through social structures and makes contributions that are valuable, although not wholly distinctive from that of economics (Granovetter 2002).

---

[4] These issues are discussed in the analytical sociology field. See for instance: Manzo (2012), Edling and Rydgren (2014).

Filippo Barbera and Nicola Negri - 9781789906851

On the other hand, the structural complexity highlighted by Sørensen suggests that the standard RCT model is not apt to grasp how socio-structural factors and processes *actually* work. RCT might certainly have the virtue of a "paradigmatic privilege" (Goldhorpe 2000; Abell 2003): it is an apt tool to measure the distance between empirical settings and those that would be given if agents would have decided according to the standard rules of the RCT in competitive environments (Coleman 1990, pp. 706–9). But, precisely this "paradigmatic privilege" of RCT highlights its key shortcoming: the principle of decreasing abstraction according to which the models must be "as simple as possible and as complex as *necessary*" (Lindenberg 1992), must be taken seriously, also in the second part of the sentence to make sense of social structure in accordance with analytical realism.

It should be stressed, moreover, that these weaknesses of RCT arise in contexts that are quite faraway from "ontological uncertainty" (Lane and Maxfield 2005), such as those contexts evoked by the most drastic criticisms of RCT (Denzin 1990). The context of the fordist assembly line is indeed quite a transparent one, and the analyses of both Sørensen and Stewman, as we have just been illustrating, point out how these contexts are easily "parameterized". The contexts in which decision-makers act are not intrinsically ambiguous or ontologically uncertain, but only *complicated*, and a wise use of modeling can make them more transparent. Nonetheless, assuming that in these contexts the micro-level mechanisms are those of RCT seems highly inappropriate with respect to the requirements of analytical realism. No less out of place than assuming that, since the game of billiards is a parametric environment (at least more than the poker one) the player calculates their side shots using the laws of physics and the rules of trigonometry. In fact, we know that billiard players proceed in a heuristic, and not computational, manner.

To better balance out parsimony and complexity, do we have to consider micro-level models that are closer to the bounded rationality approach? May we assume, for example, that in the face of computational complexity, agents apply models of bounded rationality *à la* Simon? The bounded rationality approach in which vacancies are formalized as first-order Markov processes where a decision about moving along the career ladder evaluates *only* the next vacant position (Chase 1991, p. 138) are no more compelling. Real-world agent's think of themselves in the *future* (Appadurai 2004) and their projects include *long-term* career perspectives. If individuals would not think about themselves in the future, career opportunities would lose their effectiveness and careers would no longer work as a tool for regulating organizational environments. Organizational careers, thus, imply logics of action where agents have long-term time horizons, but these cannot be understood through RCT models.[5]

1.1 Coupled Contexts and Social Rationality

At this stage of the argument, we could be tempted to ask which kind of (more realistic) micro-level models the social sciences toolbox may provide. From the examples we have just illustrated, this is not the key question to ask and, indeed, the question is badly asked. The

[5] Experimental/behavioral economics shows that rewards that are received sooner are often preferred over future rewards. RCT incentives, thus, would hardly work for long-term career time horizons. Therefore, to effectively support one's career, the decision-maker has to: "voluntarily postpone impulsive urges for immediate gratification and persist in goal directed behavior to achieve positive outcomes that are delayed in future" (Wittmann and Paulus 2009, p. 1).

Filippo Barbera and Nicola Negri - 9781789906851

point at stake is not to correct Coleman's advice by suggesting an alternative model to the RCT one on micro-level grounds per se. Rather, the theory-building process would be better to follow an alternative path which, in some ways, turns Coleman's recommendation on its head while *keeping* its foundational aim. This is the challenge that Mark Granovetter (2002) proposed for a unified social science.

> The most daunting agenda for a unified social science is to integrate such analyses (RCT analysis) with the more contextually complex arguments of structural sociology. RCT is a rather special case where context stands still and is *decoupled* from rational action in a clearly identified social space […]. The challenge for the new century is to build theory for the more general case, where *contexts*, structures and *individual actions* interact and change *together*. (p. 54, emphasis added)

It is by accepting the challenge that it is possible to deal with the mechanisms detected by Burawoy's ethnographic account, providing as well a non-RCT explanation of workers consensus and commitment in "middle-level" organized settings. In this vein, in the following we will see how organizational hierarchies regulated through job careers call for micro-level models more complex from those envisaged by the RCT. In other words, micro-level complexity and structural complexity can and should grow *together*: to grasp the complexity of social structure, a "complex enough" micro-level model is key. Coleman's suggestion has remarkably narrow scope conditions: it applies when contexts and agents are *decoupled*. But when contexts, structures and individual actions interact and change together, this suggestion is misleading.

At the micro-level, this standpoint opens toward the centrality of social rationality, while the approach mapped out by Coleman lies within the boundaries of natural rationality. While in the natural rationality approach people are rational in and of themselves and institutions are instruments for improving "collective rationality": "(i)n the other camp, scholars assume that individual rationality is something that is strongly affected by social processes. […] Social processes can increase or decrease individual and collective rationality and they can foster *different* kinds of rationality" (Lindenberg 2003, p. 2, emphasis added). This path should be taken *cum grano salis*. The pitfalls of hyper-description associated with social rationality are well-known: if the logic of action is itself a function of the context, it is very likely that the social scientist's task ends with *thick descriptions* of contexts, robbing the theory of all analytical power and, eventually, resorting to historical explanation (Barbera and Negri 2014). To avoid this hazard, the analysis would benefit from the use of a parsimonious framework, like the DBO one.

As a theory (which it is not, as we will spell out in the following), the DBO has been judged as both *too closely linked* to the rational choice model, and as being *not linked closely enough* (Pickel 2006, p. 174; Opp 2007, pp. 119–20). Thus, it has been argued that: (i) DBO theory uses an intentionalist folk psychology that is unable to account for the complexity of habitual actions, which are preconscious and rules guided (Gross 2009; see Manzo 2010 for a reply); and (ii) DBO theory is nothing more than an implicit version, albeit a "weak" one, of rational choice theory (Abbott 2007). We assume that these criticisms share a mistake, since they both fail to realize that the focus of DBO is not a theory of individual choice per se. Rather, it identifies a framework (Ostrom 1997) holding analytical elements whose varying combination with socio-structural conditions generate models of choice that come close, or fairly close, to the RCT, as well as complex interaction patterns mediated by feelings, emotions and perceptions such as a liking for homophily, or for power as authority and influence.

Filippo Barbera and Nicola Negri - 9781789906851

The shift from the field of natural rationality to that of social rationality entails a second challenge that needs to be addressed. Sociological concepts hold a peculiar quality: they cannot be both univocally defined *and* widespread in the scientific community. The more univocal their definition, the less diffuse they will be. Conversely, if a concept is widely used, its definition will inevitably be highly polysemic. Since the concept of social structure has been one of the most popular concepts in the sociological debate and is widely used in many different approaches, its polysemicity should not come as a surprise. Sociologists interpret it according to their favorite theoretical, epistemological and methodological framework; with definitions ranging from those centering on population attributes (Blau 1994), on positional considerations (Sørensen 1998) and on social networks (Burt 1992). It is far beyond the scope of this paper to systematically review the many different meanings of social structure (Porpora 1989). However, it suffices to say that the polysemic slant of the concept can hinder its analytical power when used to make sense of co-variation of contexts, structures and individual actions. A taste for complexity and for *coupled contexts* might easily produce lists of plausible factors and narratives where everything is equally key for the explanation of the phenomena (Barbera, this volume). In the next paragraph, we will deal with this tricky issue by presenting a scheme in which the parameters of the context of interaction depend on the interplay of two definite dimensions of the social structure: (i) properties of positions in formal hierarchies and (ii) properties of social networks.

2. KINDS OF RELATIONS BETWEEN PROPERTIES OF POSITIONS AND PROPERTIES OF SOCIAL NETWORKS

Social networks have emergent structural properties that differentiate them from the sum of N dyadic exchanges, unlike, for example, the market system (Granovetter 2002, 2005). Social networks affect key outcomes and phenomena and, in this regard, sociology has made major contributions to understand their effect in labor markets, organizations, business groups and innovation systems (e.g. Granovetter 2017), just to name a few. On their part, social positions have structural properties such as material and symbolic advantages, rights to act and power to command. These properties are independent of individuals characteristics, and thus "exist prior to the interaction of individuals filling these positions" (Udehn 2001, p. 305), creating effects independently of the characteristics of the people occupying them. For instance, wages and rewards in a job are not determined only by the incumbent's individual resources and "human capital" but are attached to titles or positions (Petersen 2004). Social positions therefore have structural properties, such as rent (Sørensen 1996), that generate interests in the protection and the acquisition of advantages attached to them. We will now consider several potential links between the properties of social networks and the properties of the positions in hierarchically organized systems (arrows a, b, c, d, Figure 13.1), showing how they call for micro-level complexity.

2.1 From Network Centrality to Hierarchies

Examples of relationship (a) emerge when considering the properties of highly central nodes in relatively *decentralized* networks. Occupying a central position in a "star" network—or brokering a structural hole (Stovel and Shaw 2012)—means to have access to disconnected

Filippo Barbera and Nicola Negri - 9781789906851

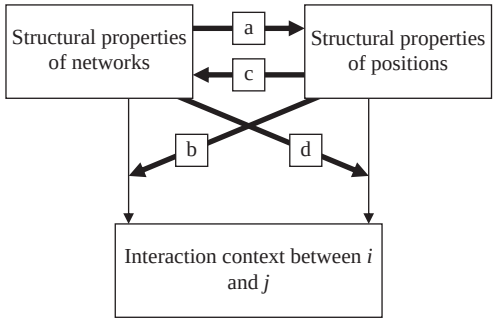


Figure 13.1 Positions, social networks and power

pieces of information, or to make gains by connecting opportunities that would otherwise be separated (Burt 1992; Padgett and Ansell 1993). Consequently, differences in the centrality generate asymmetries and uneven distributions of information and resources. In this way, differences in the centrality also create favorable conditions for setting up hierarchies.

Two generative mechanisms may be at work in this process. First, information asymmetries are an incentive to enter relational contracts (Milgrom and Roberts 1992), for these can solve coordination problems by assigning the power to make decisions to a center (Grandori 1999, p. 159). Second, the presence of powerful nodes generates asymmetries in the risk propensity of actors. The wealthier actors will differentiate more of their investments, and will be more risk neutral than the poorer ones whose income chiefly depends on their work and "human capital". In such a context, the more risk-neutral actors will be willing to purchase labor and rights to control from the more risk-adverse ones in order to take advantage of opportunities. The diversity of risk attitudes associated with the nodes centrality lays the foundations for relational contracts, governed by norms that establish controls and sanctions, whereby one party give up decision-making power for security, and the other guarantees security in exchange for power. Parsimonious RCT models may apply in this case, assuming that power does not rest on legitimacy, but rather flows from what Weber called the "constellation of interests" (1921, chapter 10), for him the least interesting cases in the analysis of power (Granovetter 2002). However, we could assume—on the shoulders of Burawoy's analysis previously recalled—that in the context of organizations governed by relational contracts, agents get their commitment both from other's respect and from their positioning in prestige/status hierarchies. It follows that the understanding of the relationships between these agents and their interaction patterns requires to step away from parsimonious explanations based on simple and pre-existing "constellation of interests" (Gould 2002; Manzo and Baldassarri 2015). We will now dig deeper in these issues, while analyzing the relationships specified by the arrows b, c and d in Figure 13.1.

2.2 Weakening and Empowering of the Power of Central Nodes

Nodes located in central positions and that in turn are in contact with other central nodes reinforce their power *if* the network is part of a hierarchical organization with transitive power relationships. This example brings us to the ambiguity existing between the centrality of a node and that of the other connected actors (see Degenne and Forsé 1994, p. 161). The

Filippo Barbera and Nicola Negri - 9781789906851

problem lies in establishing whether the centrality of an actor is strengthened *or* weakened by the fact that the actor is connected to other actors who are, in turn, central. At first glance, the answer would seem rather straightforward. A central node is an "irreplaceable intermediary", and this is where its source of power lies. Being an intermediary/broker between powerful actors should thus be an important structural factor underlying the node's power. Nevertheless, Degenne and Forsé also point to analyses such as those by Cook et al. (1983) and by Roy and Bonacich (1985), which turn this assumption on its head. Being connected to very powerful actors might lead to a weakening of power. Consequently, being the "center of centers" may not be such a good idea. This occurrence is a matter of context and should not just depend on a formal analysis of networks. From a theoretical standpoint, the centrality-power of a node can be increased, decreased or entirely unaffected by the centrality of the nodes with which it is in contact. Degenne and Forsé—and this is the point of interest here—state that it is key to understand if the network is embedded in an organizational setting equipped with *transitive power relationships*. Or, in other words, if the network is part of an organized setting with hierarchical rules whereby if A has power over B and B has power over C, then A has power over C. In this case, a node's centrality is reinforced by the centrality of the nodes with which it is in contact (cf. Degenne and Forsé 1994, p. 163). The opposite occurs when no transitive power relations are given. In this second case, who occupies the "center of the centers" may just be a *puppet* whose decisions are actually controlled by B and C, who might be in competition or in coalition between each other. Thus, in the absence of transitive hierarchical command, the node centrality is decoupled from its correlates in terms of power, influence and prestige. In these cases, those who are placed in the "center of the centers" can undergo—as a perverse effect of their visibility—a process of "moral disqualification" being considered as a "screen" for the maneuvers of those who have the real power, or being labelled as the "scapegoat" if something goes wrong. For this reason, it is rational to step away from this position either to establish a transitive hierarchical relationship or to eliminate the middle-level centers to establish a direct relationship with their subordinates. Alternatively, exchanges of material and symbolic resources can compensate those who—albeit in a higher hierarchical position—run the risks of "moral disqualification" and "scapegoating". We will see, in the following, that when the power of a center is strengthened (and not weakened) by its relations with other centers, not everything is resolved. Even in this case, logics of action far more complex than the RCT ones are required.

2.3 Positive or Negative Connections and Time-Lagged Exchanges in Hierarchical Organizations

In hierarchical organizations with transitive power relationships, what matters is if the organizational structure supports positive or negative connections between the nodes of a given network. There is a positive connection between actors A, B and C when the exchange between A and B has positive consequences on the exchange between A and C (Degenne and Forsé 1994, p. 165), or, in other words, when the exchanges approximate a positive sum game for all the nodes involved in the pairwise exchanges. Relationships are negative when the exchange between A and B jeopardizes that between A and C, or the exchanges have a negative sum for all the nodes involved. In this case, the relationship between B and C is competitive, and A cannot help B without hindering C. Given this definition, we will now turn to the cases in Figures 13.2a and 13.2b.

Filippo Barbera and Nicola Negri - 9781789906851

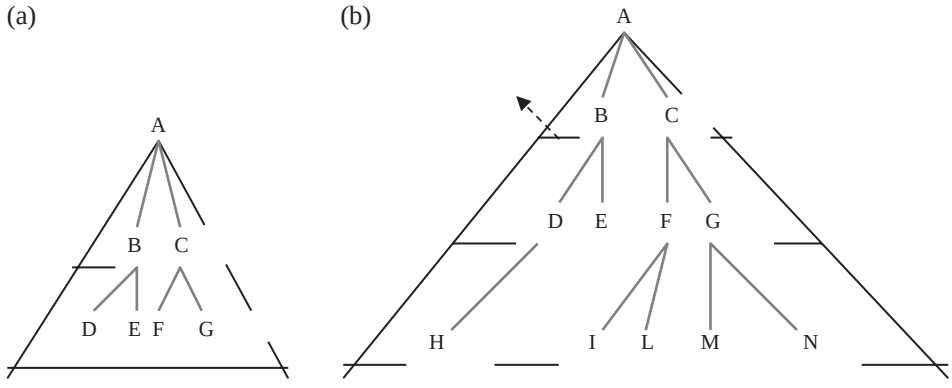


Figure 13.2 Centrality and power with negative and positive sum games

In Figure 13.2a, there are three levels: entry, intermediate and top. In Figure 13.2b, levels increase to four: entry, top and two intermediate levels. In both organizations, we assume that the personal power of those who occupy a higher position in a structure increase the more this position allows them to promote or to hire directly. The personal power also increases when the position occupied gives them the chance to promote or to hire. In all these cases, the exercise of power establishes or strengthens personal ties through which authority, influence and reputation circulates and allows to send upward signals useful for one's career.

Now, suppose in the organization depicted in Figure 13.2a the need arises for a new position at the lower level, that of the nodes D, E, F, G; at this level it is a matter of assuming a new employee. As a result, A can order B to hire a new employee. Alternatively, A could order C to do the same thing. For B and C, as we have mentioned, exercising the power to hire entails an increase in their power. A's order is then an opportunity for both B and C. In view of all this, it is straightforward that A has a competitive, zero-sum relationship with B and C: by supporting C, A withdraws support from B, and vice versa. Therefore, in structure Figure 13.2a the exercise of A's formal power may structurally jeopardize the advantages that would derive from its centrality. On the one hand, exercising its power, A risks unbalancing the relationships between the intermediate leaders by weakening the authority of one of them, with negative consequences on its own power as the "center of the centers". On the other hand, its power could also be weakened if B and C, aware of the importance of keeping a balance between them, were able to leverage their centrality to form opposing coalitions to put pressure on how to manage the power to promote/hire. In the case depicted in Figure 13.2b, on the contrary, the organizational structure gives to A the possibility to exercise their "power to promote" with advantages to *both* C and B. In this way, nonetheless, A "passes the stick" to C, who needs to deal with F and G to manage the dilemmas of a zero-sum competitive situation between them. There is a key difference: in Figure 13.2b at the top floors A's immediate chances of distributing the power to promote (direct or indirect) and the relative gains are still wider than those offered by the situation 2a. This *structural* possibility enhances the power associated with A's centrality, while the burden of being central in negative or competitive connections is passed downward and involves nodes at the lower levels of the organizational

Filippo Barbera and Nicola Negri - 9781789906851

hierarchy. At these lower levels, it is less key to take care of power equilibria, and the agents are less powerful and hence less capable to resist the exercise of power from hierarchically better positioned nodes. The greater the number of hierarchical levels, the more likely it is that positive-sum connections will extend to lower levels of the hierarchy. As a result, the ease of careers open up further possibilities for new action logics. The sheer fact that on the upper levels of the organization apical actors have enough room to maneuver is a sign of their will/ability to put into practice cooperative solutions. This might trigger positive beliefs about the cohesion of the organization *as a whole*, or about the widespread cooperation as the general working rule of the organizational setting. Agents will thus derive their expectations/beliefs on the possibility of acting their roles in the proper way and of being promoted from "signs" resulting from the situational context (see Keuschnigg and Wolbring 2015 and Barbera, this volume), quite independently from current opportunities and short-term cost/benefit computations. The organization in itself becomes the target of individual trust, and time-lagged social exchange under incomplete contractual conditions and ambiguity may flourish. According to Kanter, this is not achieved through contract-like exchanges or surveillance, but through an extreme form of the belief that individual interests are best served by a complete immersion of each individual in *the whole* (1972, p. 4). Accordingly, clans as organizational governance structures (Ouchi 1980) may grow *within* hierarchical settings. Let us now see how, in addition to time-lagged exchanges, the easing of career mechanisms also opens up opportunities for deeper changes that call into question the type-d effects (as depicted in Figure 13.1).

2.4 Intrinsic Value, Networks and Hierarchies

In a given organization, the number of hierarchical levels influences the likelihood that *extrinsic*-value exchanges generate a network whose ties have *intrinsic* value for the actors. For a given rate at which vacancies occur, the higher the number of hierarchical levels is and the larger the amount of people who will candidate for promotion, have been promoted or have decided to promote someone else will be, the higher the frequency of career-focused interactions and correlated emotional states. Here career-focused interactions will be vehicles for sociable feelings, pleasure for homophily or a carrier for reciprocal recognition mechanisms. In this kind of organization, the interactions that constitute a career path will approximate that of a *ritual* that inspires feelings of sociability and bestows an intrinsic value on these interactions, their outcomes, the relational settings in which they take place as well as on the rules that govern them (Collins 2004). Career episodes may thus acquire an intrinsic value, which is key for the social identity as in the case of moral careers (Goffman 1959). Hence, if the hierarchical structure is suitably stratified, a bunch of interactions triggered by the availability of vacancies for career advancements might generate a more complex and multidimensional organization. Here, occupying "positions" involves not only peculiar tasks, salaries and conditions of power, but also sharing a collective identity with fellow workers (Pizzorno 2006); accordingly, the exit option (Hirschman 1970) could easily trigger a crisis in one's biographical continuity.

Besides, the micro-level logic of action steps away from the RCT's standard models while remaining within the DBO framework. Beliefs about "what is worth of people" and related desires, as well as beliefs about "who you are", exercise a normative influence that transcends RCT scope conditions. On the other hand, it is rather clear that a RCT approach

Filippo Barbera and Nicola Negri - 9781789906851

would end up not grasping the multidimensional complexity of these organized settings. It needs to be underscored that this "logic of appropriateness" (March and Olsen 2006) will not hinder the adaptability of the organization. First, the presence of ties with intrinsic value can support a sort of "optimism of the will" in seeking solutions in those previously discussed cases structured by zero-sum exchanges and which make the regulation of the careers a daunting task. Second, the fact that career episodes acquire an intrinsic value may impact the desire to "move ahead" and to the commitment to work hard in order to acquire merit for promotion, regardless of the current presence of such opportunities for advancement. A commitment, therefore, that does not stem from RCT-like decisions concerning the expected utility of hard work. In this way, commitment is quite independent from *currently* existing opportunities. Third, ritual situations in which the actions acquire an intrinsic value may help to overcome the well-known resistance to innovation which emerges when the impact of innovation on the future states of the world is not yet meaningful in the daily experience (Centola 2018). Even enthusiastically adhering to an innovation, albeit in conditions of high uncertainty and perhaps even for this reason, can in fact be worthy of appreciation.

2.5 Positive Feedbacks Between Authority and Influence

Career ladders in hierarchical settings support mechanisms by which power to *dominate* on subordinates positions and power to *influence* within one's personal network reinforce each other. This point is rooted in the multidimensional models of social stratification discussed at length by Randall Collins (1988), drawing on analysis carried out as part of anthropological studies of group cultures such as those by Mary Douglas (1966). These multidimensional models are built on the assumption that social stratification cannot be reduced simply to the vertical dimension of power. The effects of belonging to status groups can also come into play. In the simplest cases, the effects are generated independently in the two dimensions that add to each other accordingly. In more complicated cases, the two dimensions *interact* and the effects on one dimension modifies the effects on the other. A first case of interaction occurs when the ability to be a role model in a group can be increased by the fact of giving orders, and this is another instance of link b in Figure 13.1. Besides, a second case of interaction is when the probability of being obeyed can increase or decrease as a result of the status level, and this is another case of link d in Figure 13.1. As a general standpoint, we assume that bureaucracies "oiled" by the career mechanisms are the type of context that foster these patterns of interaction. Here, rising to higher positions does not simply mean having more opportunities for giving orders and sanctions (the legitimate power of domination). People who have successful careers in these kind of settings also gain a "reputation for excellence": as they move ahead in the organization, they demonstrate that they are better than their peers at doing their job. Therefore, those with successful careers might be able to gain advantages in terms of both power and prestige and can accordingly exert a power to influence. Thus, the motivations for career competitions are strengthened. Reputational costs for not participating in them rise as well: not to compete, regardless of personal reasons, publicly displays the willingness not to expose oneself to meritocratic assessment. On the other hand, power to dominate and power to influence reinforce each other, making it possible to exercise power in authoritativeness modes (Degenne and Forsé 1994, p. 164). In this situation, obedience

Filippo Barbera and Nicola Negri - 9781789906851

will be based on influence, as when we when we comply with the prescription of a doctor even without directly checking the accuracy of the prescribed therapy (La Valle 1992, p. 79; Parson 1978). Again, we are dealing with a co-evolution of contexts, structures and logics of actions. The positions occupied in the hierarchical structure not only make it possible to distribute rewards and sanctions to subordinates, but to acquire reputation and exercise influence. The DBO *framework* is still useful to make sense of these kind of phenomena, while a RCT would not grasp them and would also impede making sense of the explanatory role of "contexts".[6]

3. CONCLUSIONS

We have been maintaining that—in "middle-level" complexity contexts—the complexity of the micro-level and the one of social structure should grow together. We have been further showing how the interdependence between networks and organizational positions might affect interaction contexts and decision-making models in ways at odds with the Coleman's assumption that a parsimonious action theory such as RCT is the best tool to grasp the complexity of social structure.

However, it needs to be emphasized that the overturning of Coleman's advice does not mean to give away the macro-level focus of sociological explanation. Quite the opposite: this analytical strategy is especially apt to explain macro-level emergent phenomena and/or radical innovations. To conclude in this regard, it is worth taking up the standpoint of Padgett and Powell (2012). Taking up the distinction between natural rationality and social rationality referred to in the first paragraph, we can assume that the field of natural rationality corresponds to what Padgett and Powell assign to methodological individualism and that of social rationality pairs to what they ascribe to social constructivism. Moving in the field of methodological individualism/natural rationality, they note: "the analyst takes as given some constitutive features of the hypothesized individual or actor (typically preferences, beliefs, and resources) and then derives aggregate or behavioral conclusion from them". Here, as we already noted, "actors create relations". In the field of social constructivism/social rationality, on the other hand, it is key to understand how "relations create actors" (Padgett and Powell 2012, p. 2). But—and here is where the matter lies—these two perspectives should not be opposed to each other as radically divergent: "[t]he difference methodological individualism and social constructivism is not for us a matter of religion; it is a matter of time scale. In the short run, all objects—physical, biological, or social—appear fixed, atomic. But in the long run, all objects evolve, emerge, transform, and disappear" (pp. 2–3). This perspective pushes to assume the "middle ground" of the *social organization* as the analytical starting point and then reach the macro level and the micro level. The Coleman boat syntax may still be helpful in this regard, providing that it encompass those mechanisms where actors and contexts change *together*.

[6] It is worth discussing if the "social rationality" approach we defend here is just another label for the "weak" rational choice theory (see Opp 1999). The point at stake is how far RCT model can be weakened and still be a distinctive approach.

Filippo Barbera and Nicola Negri - 9781789906851

REFERENCES

Abbott, A. (2007), "Mechanisms and Relations", *Sociologica*, **2**: 1–22.

Abell, P. (2003), "On the Prospects for a Unified Social Science: Economics and Sociology", *Socio-Economic Review* **1** (1), 1–26.

Appadurai, A. (2004), "*The Capacity to Aspire: Culture and the Terms of Recognition*", in V. Rao and M. Walton (eds), *Culture and Public Action*, Palo Alto, CA: Stanford University Press: 59–84.

Barbera, F. and N. Negri (2014), "Comment on Peter Abell, Teppo Felin and Nicolai J. Foss: Crypto-Rational Choice or Complex Mechanisms?" *Sociologica*, **2**, doi: 10.2383/78820.

Blau, P.M. (1974), "Parametres of Social Structure", *American Sociological Review*, **39**, 615–35.

Blau, P.M. (1994), *Structural Contexts of Opportunity*, Chicago: Chicago University Press.

Boudon, R. (1973), *L'inegalité des chances*, Paris: Armand Colin.

Boudon, R. (2014), "What is context?", *Köln Z Soziol*, **66**, 17–45

Burawoy, M. (1979), *Manufacturing Consent: Changes in the Labor Process Under Monopoly Capitalism*. Chicago: University of Chicago Press.

Burt, R.S. (1992), *Structural Holes: The Social Structure of Competition*, Cambridge, MA and London: Harvard University Press.

Centola, D. (2018), *How Behavior Spreads: The Science of Complex Contagions*, Princeton, NJ: Princeton University Press.

Chase, I.D. (1991), "Vacancy Chains", *Annual Review of Sociology*, **17**, 133–54.

Coleman, J.S. (1990), *Foundations of Social Theory*, Cambridge, MA: The Belknap Press.

Collins R. (1988), *Theoretical Sociology*, Orlando, FL: Harcourt Brace Jovanovich.

Collins. R. (2004), *Interaction Ritual Chains*, Princeton: Princeton University Press.

Cook, K., R. Emerson, R. Gillmore and T. Yamagishi (1983), "The Distribution of Power in Exchange Networks: Theory and Experimental Results", *American Journal of Sociology*, **89**, 275–305.

Degenne, A. and Forsé, M. (1994), *Les Réseaux Sociaux: une Analyse Structurale en Sociologie*, Paris, Armand Colin.

Denzin, N. (1990), "The Long Good-bye: Farewell to Rational Choice Theory", *Rationality and Society*, **2** (4), 504–7.

Douglas, M. (1966), *Purity and Danger: An Analysis of the Concepts of Pollution and Taboo*, London: Routledge.

Edling, C. and J. Rydgren (2014), "Analytical Sociology: Bringing Culture and Identity Back In", *Sociologica*, (**2**). https://doi.org/10.2383/78818.

Foa, V. (1991), *Il cavallo e la torre*, Torino: Einaudi.

Goffman, E. (1959), "The Moral Career of the Mental Patient", *Psychiatry*, **22** (2), 123–42.

Goldthorpe, J.H. (2000), *On Sociology*, Oxford: Oxford University Press.

Gould, Roger, (2002), "The Origins of Status Hierarchies: A Formal Theory and Empirical Test", *American Journal of Sociology*, **107**, 1143–78.

Grandori, A. (1999), *Organizzazione e comportamento economico*, Bologna: Il Mulino.

Granovetter, M. (2002), "*A Theoretical Agenda for Economic Sociology*", in Mauro F. Guillén, Randall Collins, Paula England and Marshall Meyer (eds), *The New Economic Sociology: Developments in an Emerging Field*, New York: Russell Sage Foundation, pp. 35–59.

Granovetter, M. (2005), "The Impact of Social Structure on Economic Outcomes", *Journal of Economic Perspectives*, **19** (1), 33–50.

Granovetter, M. (2017), *Society and Economy: Framework and Principles*, Cambridge, MA: Harvard University Press.

Gross, N. (2009), "A Pragmatist Theory of Social Mechanisms", *American Sociological Review*, **74**, 358–79.

Hechter, M. and S. Kanazawa (1997), "Sociological Rational Choice Theory", *Annual Review of Sociology*, **23**, 191–214.

Hedström, P. (2004), "Generative Models and Explanatory Research: On the Sociology of Aage B. Sørensen", *Structures, Dynamics and Mechanisms: Essays in Honor of Aage B. Sørensen, Research in Social Stratification and Mobility*, **21**, 13–25.

Hedström, P. (2005), *Dissecting the Social. On the Principles of Analytical Sociology*, Cambridge, MA: Cambridge University Press.

Hirschman, A.O. (1970), *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*, Cambridge, MA: Harvard University Press.

Kanter, R.M. (1972), *Commitment and Community*, Cambridge, MA: Harvard University Press.

Keuschnigg, M. and T. Wolbring (2015), "Disorder, Social Capital, and Norm Violation: Three field Experiments on the Broken Windows Thesis", *Rationality and Society*, **27** (1), 96–126.

La Valle, D. (1992), *Denaro, prestigio, e regolazione sociale*, Bologna: Il Mulino.

Lane, D. and R. Maxfield (2005), "Ontological Uncertainty and Innovation", *Journal of Evolutionary Economics*, **15** (1), 3–50.

Lindenberg, S. (1992), "The Method of Decreasing Abstraction", in J.S. Coleman and T. Fararo (eds), *Rational Choice Theory: Advocacy and Critique*, Newbury Park, CA: SAGE: 3–20.

Filippo Barbera and Nicola Negri - 9781789906851

Lindenberg, S. (2003), *Institutional Design and Its Discontent: Coleman's Neglect of Social Rationality*, Center for the Study of Economy and Society, Working Paper 14, Cornell University.
Manzo, G. (2010), "Analytical Sociology and its Critics", *European Journal of Sociology*, **51** (1), 129–70.
Manzo, G. (2012), "Reason-based Explanations and Analytical Sociology. A Rejoinder to Boudon", *Revue européenne des sciences sociales*, **50** (2), 35–65.
Manzo, G. and D. Baldassarri (2015), "Heuristics, Interactions, and Status Hierarchies: An Agent-based Model of Deference Exchange", *Sociological Methods and Research*, **44** (2), 329–87.
March, J.G. and J.P. Olsen (2006), "The Logic of Appropriateness", in M. Rein, M. Moran and R.E. Goodin (eds), *Handbook of Public Policy*, Oxford: Oxford University Press, pp. 689–708.
Milgrom, P. and J. Roberts (1992), *Economics, Organization and Management*, Englewood Cliffs, NJ: Prentice Hall.
Opp, K.-D. (1999), "Contending Conceptions of the Theory of Rational Action", *Journal of Theoretical Politics*, **11**, 171–202.
Opp, K.-D. (2007), "Book Review of Peter Hedström: Dissecting the Social. On the Principles of Analytical Sociology", *European Sociological Review*, **23** (1), 2007, 115–22.
Ostrom, E. (1997), "Institutional Rational Choice: An Assessment of the IAD framework", in P. Sabatier (ed.), *Theories of the Policy Process*, Boulder, CO: Westview Press.
Ouchi, W.G. (1980), "Markets, Bureaucracies, and Clans", *Administrative Science Quarterly*, **25** (1), 129–41.
Padgett, J.F. and C.K. Ansell (1993), "Robust Action and the Rise of the Medici, 1400–1434", *American Journal of Sociology*, **98** (6), 1259–319.
Padgett, J.F. and W.W. Powell (2012), *The Emergence of Organizations and Markets*, Princeton, NJ: Princeton University Press.
Parsons, T. (1978), *Action Theory and Human Condition*, New York: The Free Press.
Petersen, T. (2004), "Aage Bøttger Sørensen's Contributions to Quantitative Sociology: From Studying Processes to Studying Structure", *Structures, Dynamics and Mechanisms: Essays in Honor of Aage B. Sørensen, Research in Social Stratification and Mobility*, **21**, 27–44.
Pickel, A. (2006), "Book Review of Peter Hedström: Dissecting the Social. On the Principles of Analytical Sociology", *Canadian Journal of Sociology Online*, http://www.cjsonline.ca/reviews/dissectsocial.html
Pizzorno, A. (2006), "Rational Choice", in S. Turner and M. Risjord (eds), *Handbook of the Philosophy of Science, Volume 15: Philosophy of Anthropology and Sociology*, Dordrecht: Elsevier), pp. 373–95.
Porpora, D. (1989), "Four Concepts of Social Structure", *Journal for the Theory of Social Behaviour*, **19** (2), 195–211.
Raub, W. and J. Weesie (1990), "Reputation and Efficiency in Social Interactions: An Example of Network Effects", *American Journal of Sociology*, **96** (3), 626–54.
Raub, W., V. Buskens and Marcel, A.L.M. Van Assen (2011), "Micro-Macro Links and Microfoundations in Sociology", *Journal of Mathematical Sociology*, **35** (1–3), 1–25.
Roy, W. and P. Bonacich (1985), "Centrality and Power in a Balcanized Network: Interlocking Directorates among American Railroads 1886–1985", Communication presented at the Annual meeting of the American Sociological Association, Washington, DC.
Sørensen, A.B. (1977), "The Structure of Inequality and the Process of Attainment", *American Sociological Review*, **42** (6), 965–78.
Sørensen, A.B. (1996), "The Structural Basis of Social Inequality", *American Journal of Sociology*, **101** (5), 1333–65.
Sørensen, A.B. (1998), *Theoretical Mechanisms and the Empirical Study of Social Processes*, in P. Hedström and R. Swedberg (eds), *Social Mechanisms*, Cambridge: Cambridge University Press: pp. 238–66.
Sørensen, A.B. (1999), "Statistical Models and Mechanisms of Social Processes", Manuscript, Department of Sociology, Harvard University, Cambridge, MA.
Stewman, S. (1986), "Careers and Organizational Labour Markets: Demographic Models of Organizational behaviour", *Administrative Science Quarterly*, **31** (2), 212–47.
Stewman, S. and S.L. Konda (1983), "Models of Organizational Behavior", *American Journal of Sociology*, **88**, 637–85.
Stovel, K. and L. Shaw (2012), "Brokerage", *Annual Review of Sociology*, **38** (1), 139–58.
Udehn, L. (2001), *Methodological Individualism*, London: Routledge.
Weber, M. (1921), *Economy and Society*, edited and translated by Guenther Roth and Claus Wittich (1968), New York: Bedminster Press.
Wittmann, M. and M.P. Paulus (2009), "Temporal Horizons in Decision Making", *Journal of Neuroscience, Psychology, and Economics*, **2** (1), 1–11.

Filippo Barbera and Nicola Negri - 9781789906851

14. Contingency and randomness: a modal approach

Ivan Ermakoff

INTRODUCTION

Causal arguments in social sciences cannot dispense with a positive understanding of contingency. That is, they cannot avoid objectifying the defining characteristics and properties of the mode of causality they label "contingent." Failing to do so opens the door to muddled exchanges and dialogues of the deaf regarding the etiology of the phenomena we seek to explain.

The difficulty though is that statements that invoke contingency are all over the place in terms of suggested meanings. The lexical associations that trail the predicate "contingent" range from "variable" to "indeterminate," "accidental," "nonlinear" and "exogenous" among others. In these conditions, getting oriented is a challenge. Even more troubling, these pointers do not form a prima facie coherent picture. How could "variable" be on a par with "indeterminate"? What could be the connection between "nonlinear" and "exogenous"? Given this mind-boggling and, at times, contradictory flurry of meanings, it is tempting to conclude that the maze has no exit and that a disclaimer should read: "Abandon hope all ye who enter here."

With respect to analytical sociology, the matter is of importance for at least three reasons. First, figuring out what we mean when we name a referent "contingent" serves the "quest for clarity and precision" that, although not distinctive of analytical sociology, is one of its regulative principles (Hedström 2005, pp. 3–4; Manzo 2014, pp. 10–11). Second, while proponents of the analytical sociology research program advance claims that presume a positive understanding of contingency, they leave its empirical content either pending or open to various interpretations (e.g. "the lack of determinism" in Boudon 1986, p. 181; "indeterminate triggers" in Elster 1998, pp. 46, 50; "noise" in Macy and Tsvetkova 2015, p. 30). Third, the call for analytical specificity that underlies the notion of "mechanism" ultimately pertains to the effectuation of change. Clarifying the modes whereby change is effectuated is precisely the whole point of spelling out the empirics of contingency, as I shall argue in the last section of this chapter.

To grasp the nettle and figure a way out of the semantic maze noted above, we need a method. This chapter starts by acknowledging the flurry of semantic correlates that can be culled from contingency claims (Section I). Diffuse notions usually surround themselves with a retinue of surrogates. "Contingency" is no exception. Once this list of possible meanings has been couched down, Section II proposes a way to map them out by demarcating usages: nominalist versus realist, methodological versus substantive and symptomatic versus ontological meanings. These contrasted pairs form a multi-layered classification grid. Methodological and substantive usages denote a realist understanding. Symptomatic and ontological designations are substantive in character. Distinguishing these semantic registers helps overcome blatant inconsistencies.

For the purpose of examining which ones of these categories, if any, put us on the track of an empirical understanding, Section III narrows the focus down on categories that aim to have

Ivan Ermakoff - 9781789906851

ontological meaning. A concept worthy of social scientific status should be able to disclose its empirical credentials. Otherwise, it remains open to imputations of various kinds and has the epistemic status of a Rorschach blot. For an empirical concept to be duly specified, two requirements have to be fulfilled. We should be able to trace empirically the referent in light of observable markers (traceability requirement), and we should be able to differentiate this referent from cognate phenomena that have a family resemblance (differentiation requirement) (Ermakoff 2017, p. 130).

This work of analytical and empirical specification brings into relief two basic modes of indeterminacy (Sections IV and V). One is the causal indeterminacy of coincidences, which for reasons that will become clearer in Section IV this chapter designates as "Cournot contingency." The other is the indeterminacy of open-ended conjunctures ("open contingency") (Section V). Cournot contingency rests on a clause of causal independence that encompasses chance happenings in the singular (happenstance) and the plural (random variation). Open contingency takes shape through different configurations. This chapter considers three of them: (1) "possible deflagrations" characterize situations susceptible to sudden, collective and chain-like shifts in behavior ("cascades") on multiple scales; (2) "evolutionary indeterminate conjunctures" designate conjunctures in which different path-dependent dynamics can take shape; and (3) situations of "collective irresolution" are open to contrasted group stances.

Since one of the motivations for being analytically specific is to enhance our ability to ground empirically the concepts we use and, in so doing, expose them to refutation, Section VI addresses the issue of how we can validate, or refute, contingency claims. Unsurprisingly, the task is different depending on the type of contingency we are considering. We gauge the contingency of a single happening by examining the extent to which the sequences that make up its occurrence are causally independent. In a congruent fashion, we assess randomness by examining the causal independence of multiple occurrences. Gauging the contingency of open-ended conjunctures requires documenting the processes that make these conjunctures indeterminate with regard to their outcome: the emergence of threshold distributions susceptible to chain reactions (possible deflagrations), random interactions in a context of positive feedback processes (evolutionary indeterminate conjunctures) and the emergence of a decisional challenge (collective irresolution).

Armed with these leads, the last section examines how a modal understanding of contingency, that is, a conception of contingency as comprising different modes of indeterminacy, helps us investigate social change. The focus is on three venues whereby contingency makes its way in the realm of socio-historical processes: incidental agency, random interactions, and critical challenges. "Incidental agency" points to the collective impact of contingent alterations of individual action capacities. "Random interactions" affect collective dynamics in different ways depending on whether they mold initial conditions, trigger chain reactions, or induce a shift in interactional patterns. "Critical challenges" draw attention to the bracketing of structural determinations and the possibility of breaks in causal patterns.

In the course of probing the empirical meaning of contingency, the following considerations will draw on three interrelated concepts: action system, sequence and conjuncture. An action system describes how a set of actors engage one another through their action. Delineating an action system thus implies identifying (1) which actors relate to the same decisional stake and (2) the configuration of relations through which they engage one another (e.g. conflictual, competitive cooperative or hierarchical). The notion of an action system is broader than the notion of a field insofar as it does not exclusively presume that actors are engaged in competitive

Ivan Ermakoff - 9781789906851

relationships and that they share an understanding of rules and norms. A sequence designates the successive moments punctuating the evolution of an action system. A "conjuncture" captures how the members of a group experience the action system in which they take part. This experience can be heterogeneous. It is nonetheless presumed to be temporally uniform for each group member. A shift in subjective experience for some or all group members means a shift in conjuncture.

1. SEMANTIC MOTLEY

Statements of contingency abound. Most are assertoric: they contend themselves to asserting contingency without further ado. The assertion suffers no contest and grants itself *prima facie* validity. Some statements sketch contingency by reference to what it is not. In their grand scheme of things the notion is residual. In contra distinction with these two classes of statements, the repertoire of meanings presented below draws on formulations from which *volens nolens* we can infer some semantic imputation either because these formulations point to presumed semantic correlates—they bet so to speak on a meaning taking shape through lexical associations—or because they harness the notion to an explicit definition. The purpose of this repertoire is to survey the range of meanings that contingency statements have grafted upon the notion. As shorthand, I designate each semantic category by the adjective that best captures its meaning given the quotes at hand.[1]

Accidental/Fortuitous. "If Napoleon at an early stage in his career had been killed by a meteorite, that would have been the purest of pure contingencies" (Bury 1930, p. 67). Contingency in this case denotes the outcome of chance (Brucker 2001, p. 1). The crux of the parable of Cleopatra's nose is that "history is, by and large, a chapter of accidents, a series of events determined by chance coincidences and attributable only to the most casual cases" (Carr 1963, p. 128).

Agentic. "History is filled with turning points which have resolved around the life, sanity or decisions of one or two individuals" (Leff 1969, p. 52). "[The] element of indeterminacy in human conduct makes contingency inseparable from history" (ibid., p. 54). Key to this conception is the reference to the role played by some individual actors: through their actions, failures, lapses or absences, these individuals crucially shape historical developments. "Contingent happenings are subject to the vagaries of human will" (Sewell 2005, p. 197).

Conditional. "The event's exact unfolding … is contingent on the precise actions constituting the event" (Griffin and Ragin 1994, p. 16). Contingency implies dependence on antecedents. Along these lines Coleman (1966) defines "the problem of contingency" as the problem of figuring out whether rational and self-interested actors will make their actions "contingent on the participation of others" (p. 616).

Conjunctural. "In sharp contrast to critical realism, theoretical realism deploys its causal entities within general covering laws and abhors conjunctural causation or contingency" (Steinmetz 1998, p. 173). "Contingency here means that complex events are codetermined

[1] While this repertoire aims to cover as much ground as possible, aiming for an exhaustive list would most likely be a vain enterprise given the looseness of the categories mentioned here and there.

Ivan Ermakoff - 9781789906851

by constellations of causal mechanisms" (ibid., p. 177). This understanding goes against a conception of causality in terms of law-like generalizations.

Correlated. "I term any measure of the total deviation of the classification from independent probability a measure of its contingency. Clearly, the greater the contingency, the greater must be the amount of association or of correlations between the two attributes" (Pearson 1904, pp. 5–6). Contingency denotes a relationship of numerical association indicating a departure from the assumption of probability independence.

Discrepant amplitudes. "[Contingency] is a feature of political life in settings where self-reinforcing processes are at work. Relatively small events, if occurring at the right moment, can have large and enduring consequences" (Pierson 2004, p. 44). Undergirding this meaning is an observation about a difference in between cause and effects. Small happenings (e.g. individual moves, accidents, lapses, misunderstandings) have big repercussions.

Exogenous. In "moments of contingency … unpredictable or exogenous events channel history in new directions" (Haydu 2010, p. 31). Contingency designates what is outside a domain of empirical investigation. For instance, Haydu (2010) refers to "war and political mayhem" as "contingent sources of revolution" (p. 37).

Indeterminate. "The event is a contingent happening because it did not have to happen as it did" (Griffin and Ragin 1994, p. 16). The contingent designates the property of what could not have been or what could have been otherwise. For instance, "in Event Structure Analysis, events are conceptualized as causally indeterminate contingent happenings, in that they did not have to unfold as they did" (Griffin and Korstad 1995, p. 449).

Nonlinear. "Temporal process models" [need to be] "sensitive to forms of historically contingent change in the form of nonlinear relations" (Isaac et al. 1994, p. 139). Accordingly, the analysis of contingency focuses on "nonlinear historical process through time" (ibid., p. 115).

Open-ended. "Insofar as the members of the group under challenge make their choices conditional on one another's and realize that they face the same uncertainty, the collective conjuncture is open-ended. It can go one way or another…. This analysis spells out the notion of historical contingency" (Ermakoff 2010, p. 106).

Residual. "In the actual practice of research, social analysts will consider an event to be contingent when its explanation appears to fall outside of existing scientific theory" (Mahoney 2000, p. 514). The contingent is what analysts define as unaccountable given the theory on which they rely. Ertman (1997) draws on this meaning to account for the two cases that deviate from his theory of European state formation ("In both [Sweden and Denmark] powerful contingent events conspired to confound expected paths of development", p. 33).

Unnecessary/Undirected. "Genocide is a contingent outcome, not the inevitable result of the world historical developments [Mann] emphasizes" (Straus 2007, p. 488). No inherent necessity is under way (Hawthorne 1991, p. 14). The sequence of happenings does not follow a logic that can be spelled out before its advent.

Unpredictable/Unforeseeable. "[I stressed] the paramount importance to history of … contingent, essentially unforeseeable events—whether of the sort we commonly attribute to chance or of the sort that displays genuine originality, inventiveness, initiative" (Gallie 1964, p. 103).

Variable. "Contingency and variable are but different terms for the compositional elements of an emerging event" (Porter 1981, p. 39). "Contingent determination appears as historical variability in a parameter regime displaying a variety of possible patterns" (Isaac et al. 1994, p. 116).

Ivan Ermakoff - 9781789906851

2. CLASSIFICATION GRID

This survey of possible meanings is at once perplexing and telling: semantically, "contingency" is all over the place. Polysemy prevails. Equally perplexing: there is no obvious thread, no apparent semantic pattern, connecting the dots. These have a composite character. Some categories play a consonant tune (e.g. "conditioned," "variable" and "conjunctural"). Others have no apparent connection. "Conditional" and "indeterminate" for instance form an odd couple since the relationship between an effect and the antecedent that conditions this effect can be determinative. If so, it is unclear how "conditional" could go along with "indeterminate." Still other categories appear to be incompatible like magnets repelling one another (e.g. "indeterminate" and "correlated").[2]

Can we discern some logic underneath the jumble? To order the tapestry (or the jigsaw puzzle, depending on the metaphor of choice), I consider three dichotomies. The first contrasts a nominalist representation, which portrays contingency as the residual of a theory or a framework, with a realist conception that assumes characteristics independent of theory. Focusing on realist conceptions, the second dichotomy draws a boundary between a methodological and a substantive understanding. The third dichotomy sorts out statements that aim to provide substantive meaning into two classes: symptomatic and ontological (Figure 14.1).

2.1 Nominal versus Real

A nominalist usage of contingency apprehends it as residual and relative. Cast in this mold, the notion basically designates what remains unexplained. See for instance the following statement: "a contingent event is therefore an occurrence that was not expected to take place given certain theoretical understandings of how causal processes work" (Mahoney 2000, p. 513). This approach to contingency suffers from two redhibitory defects. First, since it makes the content of the notion relative to the choice of a "certain theoretical understanding," it also makes this content variable. What is contingent by reference to one theory may not be by reference to another. Down the line, the meaning of contingency is evanescent. "Although the outcome is potentially consistent with the predictions of other theories not examined, the analyst deemed it to be contingent because its occurrence directly challenges the specific theoretical framework of interest" (ibid., p. 514). Vagueness is the order of the day.

Second, as nominalist usages are put to work, they reveal themselves to be incoherent. When Mahoney (2000) states: "institutions … are especially capable of seizing opportunities provided by contingent events" (p. 515) or "contingent events select institution" (p. 519), he assumes a class of events that can be deemed "contingent" because of the properties they display not because of the theory taken as reference. This implicit realist understanding in nominalist usages is not an occasional lapse. See for instance: "path dependence … involves both tracing a given outcome back to a particular set of historical events, and showing how those events are themselves contingent occurrences that cannot be explained on the basis of prior historical conditions" (Mahoney 2000, pp. 507–8). To say that "contingent occur-

[2] The present diagnosis is in line with Schedler's (2007) claim that "contingency inhabits a somewhat disorderly semantic field" (p. 54) and Ballinger's (2013, pp. 48–50) remarks about the confusions generated by the multiplicity of meanings attached to the notion.

Ivan Ermakoff - 9781789906851

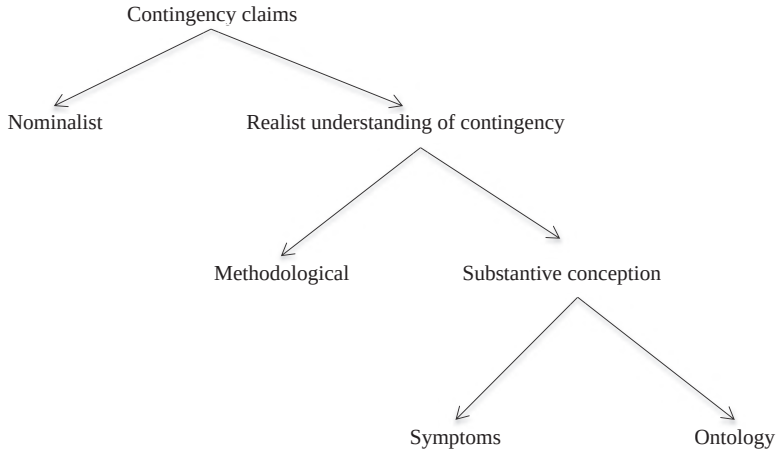


Figure 14.1 Contingency claims: epistemic registers

rences … cannot be explained on the basis of prior historical conditions" makes the predicate "contingent" independent of the choice of a theoretical framework. The imputation is a realist one.[3]

We should therefore beware of nominalist usages and categories that lend themselves to an interpretation in terms of residuals. The repertoire outlined in Section I lists three categories that run that risk: "conjunctural, exogenous and unpredictable." There is always the possibility of labeling "conjunctural" or "exogenous" factors and events that a theory does not predict, and to invoke this predicate as another name for ignorance and imperfect knowledge. Similarly, "unpredictable" easily falls on the side of a nominalist understanding. Claims of unpredictability then become relative.

If contingency is to have any positive meaning, we should, therefore, aim at harnessing "conjunctural, exogenous and unpredictable" to a realist understanding of contingency, that is, an understanding premised on what could be called a clause of epistemic independence: the empirical content of the category is independent of the conceptual act that states it (Grafstein 1988, pp. 12–14). "To be a realist at all [means] assign[ing] mind-independent status to elements of the world" (Somers 1998, p. 744).

2.2 Methodological versus Substantive

Among the realist usages of "contingency," a distinction can be made between methodological and substantive statements. Methodological statements use the predicate "contingent" to denote a relationship of association. This usage—embodied in the "correlated" category—draws attention to a type of variation that might reveal a relation of dependence. Pearson (1904, p. 5) makes such an understanding explicit when for the purpose of analyzing

[3] Also surprising is the looseness with which the article outlines the theoretical yardstick by reference to which an outcome can be labeled contingent: "theory" (p. 513), "certain theoretical understanding" (p. 513), "one or more explanatory theories" (p. 513), "existing scientific theory" (p. 514), and "prevailing social theories" (p. 514).

Ivan Ermakoff - 9781789906851

cross-tabulated data, he defines contingency as deviation from the assumption of probability independence.

While methodological statements uncover a relationship of association among variables, substantive statements refer to a mode of causality. Events deemed contingent in a substantive sense are assumed to come about in a way that belies the assumptions built in models of causal invariants and uniformities. Substantive and methodological references to contingency are at loggerheads. The methodological usage starts with deviations from the independence assumption in order to assess a possible pattern of association among variables. The purpose of this assessment is to probe and specify an explanatory model. By contrast, contingency statements understood in a substantive sense presume the irrelevance of such models. They state, posit, or conclude that we cannot predict outcomes on the basis of the patterns of association that the methodological conception labors.

The methodological/substantive distinction proves helpful when we seek to clarify the epistemic status of the "conditional" and "variable" categories. Taken at face value, these are ubiquitous categories. We would be hard-pressed to identify a social or historical process that is *not* at some stage conditional on another. The same remark holds for variation: a lack of variation is but the artefact of data censorship. Interpreted in a substantive sense, these categories are devoid of content. Being ubiquitous, they add nothing to our understanding of contingency as a predicate of events, processes or conjunctures. We get around this elusiveness if we anchor "conditional" and "variable" in the methodological camp (Figure 14.2). This qualification has the additional advantage of resolving the inconsistency noted between "conditional" and "indeterminate." If we interpret "conditional" in a methodological sense

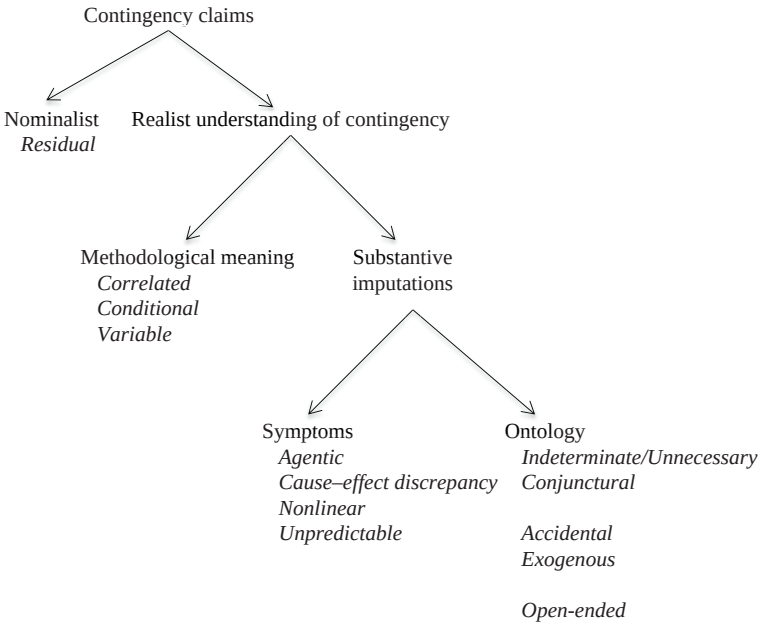


Figure 14.2 Categories of meaning sorted out by epistemic registers (the semantic correlates listed in Section 1 are in italics)

Ivan Ermakoff - 9781789906851

and "indeterminate" in a substantive one, these two categories belong to different epistemic registers. As such, they no longer contradict each other.

2.3 Symptomatic

Our focus will therefore be on statements and categories that purport to have substantive meaning. We cannot hope to substantiate the empirical content of contingency unless we investigate semantic affiliates that aim to shed light on properties from a realist standpoint. Having made this point, it does not take long to realize that several substantive categories ("agentic, discrepant, nonlinear and unpredictable") are compatible with a deterministic reading as much as they are with claims of indeterminacy. These categories are thus semantically ambiguous.

Consider the reference to individual agency. More often than not, individual actions fit highly predictable patterns and give credence to the hypotheses that they are causally driven. Instances of that kind belie Leff's (1969, p. 54) and Hawthorne's (1991, p. 14) assertion that human agency is fundamentally indeterminate.[4] The matter is actually undecidable. Reflecting on the ways in which "individual ends are formed and pursued," Goldthorpe (2016) observes that "they are socioculturally structured to some extent" and yet may represent "the ultimate indeterminism in human social life" (p. 25).

Congruent remarks apply to semantic correlates such as "cause-effect discrepancy" and "nonlinear." Both notions are key for instance to chaos theory: the chaotic behavior of a system displays "acute sensitive dependence" on initial conditions and nonlinearity (Richards 1997, p. 111; Brown 1997, pp. 53–4). Yet, as formal theorists of chaotic processes have underscored, nonlinearity does *not* imply a lack of deterministic processes (Thom 1983, p. 16; Ruelle 1991, p. 45). Nor does the fact of being unpredictable, or unforeseeable, imply indeterminacy (Grafstein 1988, pp. 12, 14; Fearon 1996, p. 42 citing Wolfram 1984, p. 32).

Given these ambiguities, what shall we make of these categories? Obviously, they cannot pretend to capture intrinsic properties. That is, they cannot pretend to have ontological meaning. On the other hand, they may have the significance of conditional symptoms. For instance, a single action's ability to generate large-scale behavioral changes can be symptomatic of relational configurations liable to chance happenings. Changes in the "magnitudes of causal relations" across time may indicate breaks in causal patterns (Isaac and Griffin 1989, pp. 879–80; Isaac et al. 1994, p. 116).[5] Unpredictability for its part may reflect the cumulative impact of random factors.

3. ONTOLOGICAL CLAIMS

Our primary focus should therefore be on six categories: "accidental, conjunctural, exogenous, indeterminate, open-ended, and unnecessary" (Figure 14.2). Do these categories help us identify contingency as "a property peculiar to certain processes" (Boudon 1986, p. 181)? From a

---

[4] On this point, Spinoza's disclaimer regarding the notion of "free will" comes to mind: we are all the more likely to view ourselves as free when we are conscious of our desires and ignorant of the causes that determine us (Letter LVIII to Schuler, no date) (Spinoza 1966, p. 302).

[5] "The [backward moving covariance] procedure, used to assess the temporal stability of time-series estimates and to establish "shift" or "break" points in series, persuasively shows the historically contingent nature of these relationships" (Isaac and Griffin 1989, p. 879).

Ivan Ermakoff - 9781789906851

realist and ontological standpoint, we designate as contingent what happened but could have not happened, as well as what did not happen but could have happened (Bouveresse 1993, p. 127).[6] In so doing we invoke a mode of causality, that is, a way whereby outcomes and processes come about or get produced, or fail to get produced. And we place at the helm of this mode the commandeering and formidable claim of indeterminacy: contingent phenomena are *somehow* indeterminate. The challenge lies in delineating what lurks behind this *somehow*. Among the six substantive categories just mentioned, "indeterminate" and "unnecessary," which are interchangeable, have the status of conceptual explananda and question marks. These are the categories in need of empirical specification.

"Conjunctural" hardly helps us on this front. Causation is deemed "conjunctural" when it involves unique and variable "constellations" of factors" (Steinmetz 1998, p. 177).[7] This definition of contingency is wanting in two respects. First, "conjunctural" lacks the specificity required to differentiate a mode of causality characterized by its indeterminacy. By virtue of being social, causation is bound to be "conjunctural" if only because any social process involves actors exposed to a host of different causal influences. Second, a "constellation" may reflect a pattern of tightly coupled factors and processes. Or it may be coincidental. Revealingly, Ragin (1987), who coined the "conjunctural causation" expression, sets forth formulations that entertain both interpretations: "conjunctural causation" encompasses conditions that happen to "coincide" (pp. 25, 43) as well as "patterns of invariance and constant association" (p. 51). The mode of causality is up in the air.

By contrast, "accidental/exogenous" and "open-ended" put us on specific empirical tracks. A happening is said to be "accidental" or "fortuitous" when a coincidence of processes and factors—exogenous to one another— brings it about (Carr 1963, p. 128). A conjuncture can be depicted as open-ended when we can show that it beholds different possible futures at once. In either case, a certain indeterminacy is at work. While the indeterminacy of accidents points to the modality of their production, the indeterminacy of open-ended conjunctures bears upon the immediate future. The next two sections further elaborate these observations.

4. INTERSECTIONS WITHOUT A CAUSE

Underlying the contingency of chance happenings (happenstance) is a generic formula, which Augustin Cournot ([1851] 1975) stated with utmost clarity: "fortuitous events are events brought about by the combination or the conjunction of other events involved in *independent* series" (p. 34, emphasis added). Happenstance resulting from the intersection of sequences that belong to different phenomenal universes (e.g. biological processes, the physics of natural phenomena or human action) are particularly apt to illustrate this definition. The example of a fly hitting the eye of a motorcyclist riding their motorcycle is a trivial yet relevant example (Hall 1999, p. 160). The causal sequence leading the fly to be on the trajectory of the motorcyclist

[6] "At a minimum," when we characterize a referent as "contingent," we are saying that [this referent] did not have to be at it is" (Shapiro and Bedi 2007, p. 1).

[7] "[Contingency] also implies that such constellations are not repeatable in a general way and also that the components that make up the causally effective constellation may vary" (Steinmetz 1998, p. 177).

Ivan Ermakoff - 9781789906851

is independent from the sequence whereby the motorcyclist decided to ride their motorcycle on that road without wearing protective glasses. These sequences being causally independent, their intersection has no cause proper. It is arbitrary.[8] And since this intersection has no cause, we are entitled to describe it as causally indeterminate.

The example of the fly and the motorcyclist is trivial, discrete and singular. Cournot's empirical definition of chance can be extended to happenings that depart from these characteristics insofar as they are historically significant, have a processual character or form a series. Serendipity in science offers an eloquent illustration of "Cournot happenings" if by this expression we mean happenings that are the outcome of unconnected causal sequences. Wilhelm von Röntgen's discovery of X-rays in his laboratory at the University of Würzburg in the fall of 1895 is a case in point (Kuhn 1996, p. 57). Von Röntgen was applying high voltages to a shielded vacuum tube at night. He then noticed that each time he was proceeding with the experiment, crystals of barium platinocyanide, which were laying at some distance from the tube, glowed (Chalmers 1952, p. 218; Glasser 1959, p. 11). Had these crystals not been placed in the vicinity of the tube for reasons unrelated to the conduct of the experiment, Von Röntgen would not have realized that the voltages were producing waves that had not yet been identified.

Marshall Sahlins's (1985) analysis of Captain Cook's and his crew's interactions with Hawaii inhabitants in January–February 1779 offers a processual account of Cournot contingency. Cook anchored his vessel *Discovery* in Hawaii on 17 January 1779. The timing of his arrival coincided with the starting of the Makahiki ceremony. As a result, Hawaiians approached him as the image of the divine entity involved in the ceremony (Lono) (Sahlins 1985, p. 105). The coincidence enacted a powerful symbolic template. The *Discovery* left three weeks later as Cook's Hawaiian interlocutors were signaling that it was now time for the newcomers to leave. After a few days, unexpectedly, Cook decided to sail back to Hawaii in order to repair the defective foremast. In sharp contrast with the sentiments that their arrival had elicited a few weeks earlier, interactions between the crew and the Hawaiians were tense and antagonistic: the British's arrival was "out of phase with the Hawaiian ritual cycle." Violence and skirmishes ensued, ultimately culminating in Cook's killing on 14 February 1779. "Chance," Sahlins argues, had presided over this unexpected chain of happenings: chance understood as "the intersection of two independent chains of causation. The weak link in one of the chains was the 'dishonest work' and 'slovenly supervision' of the Deptford naval yard" [in charge of providing the foremast] (Sahlins 1985, pp. 126–7).

The notion of "randomness" generalizes the Cournot contingency to multiple occurrences of the same type (e.g. the throws of a dice). We characterize a series of throws as random insofar as we believe each throw to be the unique product of independent factors and processes. By the same token, because we believe each throw to be idiosyncratic with respect to its production, we also have good ground to presume the throws to be causally independent from one another. In these conditions, across occurrences no factor shapes factorial combinations and the likelihood of outcomes. Variation in outcomes reflects the variation of arbitrary conjunctions involving independent factors. The mode of causation is contingent in Cournot's sense.

---

[8] "There is no reason why here rather than there, why now rather than then" (Blaise Pascal [1670] 2004, p. 91).

Ivan Ermakoff - 9781789906851

5. OPEN-ENDED CONJUNCTURES

In the same way the Cournot contingency can be conceived as happenstance or as random variation, the indeterminacy proper to open-ended conjunctures can appear in different guises (Figure 14.3): (1) it transpires in action systems susceptible of undergoing abrupt and disruptive shifts in states (deflagrations); (2) it characterizes interactions on the cusp of contrasted evolutionary paths; and (3) it reflects the irresolution of groups dumbfounded by a decisional challenge. The following remarks discuss these three modes of open contingency, first by briefly delving into an empirical case: May 1917 in Saint Petersburg (regarding the possibility of deflagrations), the state of nuclear reactor technology in the 1950s (to illustrate evolutionary indeterminate conjunctures), and the meeting of the Presidium of the Serbian Communist party on 18 September 1987 (as an instance of collective irresolution). Each case sets the ground for the characterization of an empirical class and an outline of analytical bearings.

5.1 Possible Deflagration

About four weeks after the toppling of the Tsarist regime, Trotsky (1932) describes the political situation in Saint Petersburg as pervaded with considerable tension. Members of the provisional government were issuing contradictory or ambiguous statements regarding the continuation of war. Soldiers, workers and political activists were expecting the termination of Russia's military engagement. On May 3 (Julian calendar), the press published a note by the Minister of Foreign Affairs, Miliukov, signaling a renewed commitment to war. Demonstrations followed. "Crowds gathered, meetings assembled, they wrangled at street corners" (p. 355). The next day (May 4), "there were skirmishes and casualties … The encounters became more and more fierce [between the supporters of the government and the workers]; shots were interchanged, and towards afternoon they became almost continuous. Nobody knew exactly who was shooting or why, but there were already victims of this disorderly shooting, partly malicious, partly the result of panic. The temperature was reaching red heat" (p. 358).

For the purpose of delineating an empirical class, three facets of this case are noticeable: (1) the protagonists' mutual focus and emotional reactivity, (2) their readiness for "transgressive contention"—to use Biggs' (2005, p. 1686) formulation—and (3) the prospect of a "deflagration" understood as a large-scale cascade of disruptive actions. First, there is a marked increase of mutual and emotional attention within and between groups. Actors become

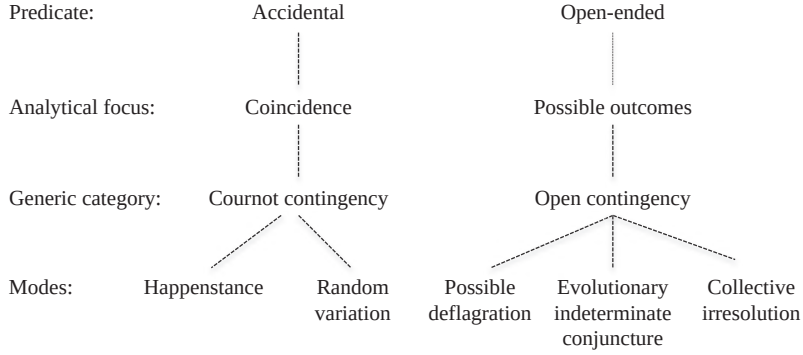


Figure 14.3 Modes of indeterminacy

Ivan Ermakoff - 9781789906851

more attentive to statements and behaviors, eager to interpret them as revealing motives, and prone to emotionally react to these interpretations. Second, protagonists are signaling their readiness to transgress the status quo. They do so through various cues: expressions of exasperation, verbal commitments, or behaviors indicating their willingness to bite the bullet (e.g. involvement in groups that flex their coercive muscles). Third, and concomitant, the situation is tense, marked by the shared realization that all hell might break loose.

Several analytical leads are available to tackle these empirical observations. Mutual focus and emotional reactivity fit arguments emphasizing the impact of physical density on emotional energy (Collins 1993, p. 206; 2004, p. 48). The second feature above—actors' readiness for "transgressive contention"—can be interpreted in terms of "action thresholds," i.e. individual actors' propensity to engage in a specific behavior depending on how many others have already done so (Granovetter 1978, p. 1422). As for the perception—pervasive among historical actors—that a deflagration might happen, this empirical observation is congruent with formal accounts devised to account for the sudden occurrence of large-scale disruptive cascades (Bak et al. 1988, p. 365; Brunk 2002, p. 210; Biggs 2005, pp. 1703–9).

5.2 Evolutionary Indeterminate Conjuncture

In the 1950s the range of possibilities regarding nuclear reactor technologies was quite wide. At the 1955 Peaceful Uses of the Atom Conference in Geneva, presenters mentioned one hundred reactor types "not obviously impractical" (Cowan 1990, p. 547). Researchers in the United Kingdom and France were working on gas graphite reactors. In Canada, the focus was on heavy water reactors. The U.S. Atomic Energy Commission (AEC) for its part was doing research on different options, including the light water reactor type. Engineers, designers and policy decision-makers were assessing the merits and demerits of these alternatives in light of numerous technical and economic parameters (Cowan 1990, pp. 545–7). Given the multiplicity of dimensions to be taken into account, no technology had a clear-cut advantage.

This case stands for a class of conjunctures in which actors have to choose between alternative ways of doing or proceeding, be these embodied in conventions, types of contract (Young 1998, chapter 1), technologies (David 1985; Cowan 1991) or locations (Arthur 1994, chapter 4). Potential adopters lack parametric information regarding the comparative merits, or disadvantages, of available options. Given this uncertainty, they have an incentive to heed the choices made by others under the assumption that these choices reveal an assessment of worth and will provide opportunities for learning. Strategic considerations thus relay parametric ones. For this class of decision problems, several equilibrium solutions in the game theoretical sense of the term are possible.[9] "But while information on preferences, endowments, and transformation possibilities allows us to locate and describe these various possible equilibria, it is usually insufficient to tell us which one will be 'selected'. There is an indeterminacy of outcome" (Arthur 1994, pp. 13, 51).

The analytical framework deployed to investigate this type of conjuncture has been centered on patterns of incentives, adaptive mechanisms and choice sequences (Arthur 1994, chapter 2; Young 1998, chapter 2). From this perspective, indeterminacy in evolutionary terms stems

[9] The notion of equilibrium is here understood in Nash's sense as designating the array of strategies such that no actor taking part in the interaction has an incentive to deviate given the other actors' strategic choices (Kreps 1990, p. 28).

Ivan Ermakoff - 9781789906851

from the combination of positive feedback mechanisms with random perturbations (Hill 1997, p. 206). Positive feedback can be of different kinds: technological externalities, learning, uncertainty reduction or increasing returns (Cowan 1991, p. 807; Cowan and Gunby 1996, p. 523; Arthur 1994, p. 111). Whichever feedback obtains, the logic of the situation is such that the more prevalent one particular option appears to be, the greater the incentive to adopt it. As a result, random factors giving precedence to one option at the expense of others can set the action system on a path-dependent evolutionary trajectory (Arthur 1994, pp. 17, 115).

5.3 Collective Irresolution

On 18 September 1987 the members of the Presidium of the central committee of the Serbian League of Communists met to deliberate on a motion set forth by the chairman of the Serbian Communist party, Slobodan Milošević, indirectly calling into question the authority and legitimacy of the head of the Serbian state, Ivan Stambolić. In effect, Milošević, acting as party chairman, was making a bid for the state leadership. Interviews and primary accounts underscore deep insecurity on the part of the Presidium members (Ermakoff and Grdesic 2019, pp. 569–73). Participants were unwilling to take a stance. Due to the unexpected character of Milošević's power bid, they did not know where the majority stood. Putting at stake their career, political status and ethics, they feared being isolated.

This event exemplifies a class of situations in which the members of a group, when confronted with a choice that they perceive as significantly risky, make their actions conditional on one another's. No behavioral script providing a way out of the challenge is readily available. Either the situation is unprecedented, or scripts inherited from the past fail to generate expectations of coordination. Therefore, actors under challenge cannot assume that those facing the same decisional stake as themselves will adopt a definite and predictable line of conduct. These are moments of *collective* irresolution. Group members acknowledge to one another their mutual dependence and their uncertainty.

The analytics of conjunctures of mutual uncertainty bring two research issues to the fore. One concerns strategies of uncertainty resolution (Ermakoff 2008, chapter 6). Individuals can resolve their decisional dilemma by aligning their behavior with one another's (sequential alignment). Alternatively, when sequential alignment does not, or cannot, take place, individuals under challenge seek to draw inferences about the behavior of the group as a whole by relying on either the information provided by interpersonal contacts (local knowledge) or the implications of public events on shared understandings (tacit alignment). The second research issue pertains to the factors that shape processes of affiliation, or identification with, a reference group (Merton 1968, pp. 284–6). Individuals under challenge vary with regard to how they configure the group they take as reference. For instance, they may refer themselves to an imagined community (e.g., "the working class"). They may fall back on the advice or injunction of close affiliates. Or they may decide to relate their decision to the stance of those directly confronted with the same decisional imperative as themselves (Ermakoff 2008, pp. 191–2).

5.4 Interactions and Indeterminacy

Whether we are considering the possibility of a deflagration, evolutionary indeterminate conjunctures, or collective indecision, three traits highlight the open-ended character of these

Ivan Ermakoff - 9781789906851

conjunctures. First, interactive processes drive the dynamic of the action system. The tension inherent to situations marked by the possibility of a deflagration stems from an increase in actors' reactivity to each other in the course of their interactions (Brunk 2001, p. 427). Evolutionary indeterminate conjunctures undergo divergent dynamics depending on the outcomes of local interactions at an initial stage (Young 1998, p. 52). In times of collective irresolution, actors probe one another and tacitly interact (Ermakoff 2008, chapter 6).

Second, uncertainty understood as a micro-analytical concept, that is, as describing individual actors' subjective experience, pervades open-ended conjunctures. When a large-scale deflagration seems imminent, actors have to decide whether to cross the line between transgressive and regulated action. In evolutionary indeterminate conjunctures, they have to appraise the risks and consequences of alternative options. When they are collectively irresolute, uncertainty pervades their perception of the situation and their view of each other.

Third, these conjunctures are indeterminate with regard to their immediate future. That is, different possible future outcomes coexist in their midst. By way of consequence—this is the fourth feature—these conjunctures are sensitive to random perturbations. As a deflagration becomes possible, an inadvertent move can generate a cascade soon experienced as irrepressible. Random interactions can give precedence to one collective scenario at the expense of others (Arthur 1994, p. 27; Hill 1997, pp. 198–9). Groups engulfed in collective indecision are susceptible to unexpected alignment initiated by lapses and mistakes (Ermakoff 2001, pp. 255–6; 2010, p. 106).

6. VALIDATION PROSPECTS

The point of crafting concepts and claims that are analytically specific and empirically grounded is in the first place to put at bay rhetorical sleights of hand, evocative poses, and the cultivation of vagueness. This should be the first payoff. There is a second one: as we identify which evidence speaks to the claims we set forth and which indicators help disprove them, we get a sense of possible validation strategies. These, as we could have expected, are functions of the mode of indeterminacy under scrutiny. Broadly speaking, probing the Cournot contingency requires assessing a premise of causal *independence*. In contra distinction with these requirements, validating open contingency requires investigating dynamics of interactional *interdependence*.

6.1 Causal Independence

The Cournot contingency can take place as happenstance (singular mode) or random variation (plural mode). In the case of happenstance, validation means assessing whether the sequences involved in the production of the happening are causally independent. Such an assessment is unlikely to be problematic when the sequences at play unfold in different phenomenal realms (e.g., a fly whose trajectory intersects a motorcyclist's). It is more arduous otherwise, especially if the sequences involve individuals partaking in the same action system, that is, orienting their action to the same stake.

Two probes are possible. One examines whether actors were enacting strategies of action without consideration for one another. If the answer is positive—actors were either unaware

Ivan Ermakoff - 9781789906851

of, or oblivious to, each other—we can presume independence. Given its focus on actors' subjective orientations, this first probe can be labeled "phenomenological." The second probe rests on a counterfactual analysis. It consists in examining whether minimally rewriting one moment in a sequence would have changed the outcome (Lebow 2000, p. 593). The higher the number of moments satisfying this counterfactual rewrite exercise, the more contingent the happening. "An outcome that requires the confluence of many independent causes but could be prevented by removing any one of them with a minimal rewrite … is highly contingent (ibid., p. 597).

As a test case, let us examine the Archduke Franz Ferdinand's assassination in Sarajevo on 28 June 1914 (Lebow 2000, p. 606). The two main protagonists were the conspirators and the Archduke's retinue. Reports concerning the threat of an attack had been conveyed to the Archduke and his associates. In the morning of the Archduke's visit on 28 June 1914, an attack did take place: one of the conspirators threw a bomb at the Archduke's carriage. The bomb bounced off and the attack failed. The Archduke's retinue subsequently decided to change the itinerary of the visit to derail the possibility of a future attack. Conspirators and the Archduke's retinue were, therefore, adapting their behaviors in light of their expectations about each other's action. The phenomenological probe mentioned above is irrelevant. Does a minimal rewrite counterfactual probe apply?

The assassination was possible because one of the conspirators, Gavrilo Prinzip, found himself close to the Archduke's carriage when it stood still. If we narrow the focus down on this moment in the sequence, three observations, consistent with a minimal rewrite criterion, give credence to the conclusion that the assassination could have easily not taken place. First, the driver of the Archduke's car had not been told of the change in the itinerary. Had he been told, he would not have turned and then been asked by one of the Archduke's aides to go into reverse. Second, as he tried to go into reverse, the car stalled and got immobilized. Had the car not been immobilized, Prinzip would not have been able to get close to it and to shoot at point blank range. Third, after the failure of the first attack, Prinzip had been wandering in this area of Sarajevo. He does not seem to have planned his presence at this intersection at this particular moment. Hence, the fact that he found himself close to the archduke's carriage at the juncture when the car stood still can be diagnosed a fluke.

In the case of random variation, validation requires certifying the causal independence of multiple occurrences. Two venues for certification are conceivable. The first examines the ways in which these occurrences have been produced or have emerged. Any indication that past occurrences affect the production of subsequent ones is enough to ruin the assumption of independence. For instance, the fact that revolutionaries often draw their inspiration from previous revolutions (see Sohrabi 2011, pp. 6–7) problematizes the assumption of causal independence that is customary in macro-comparative analyses. Randomization is a way of making sure that arbitrary conjunction—the kernel of the Cournot contingency—is the modus operandi of case selection. In the absence of randomization, or if the analysis of the processes whereby the data has been generated proves inconclusive, the second venue for certification documents the lack of evolutionary or correlational patterns. This approach can be viewed as Popperian insofar as we declare a set or series as random unless it is proved otherwise. The trouble with deterministic processes is that they can produce seemingly random outcomes (Ekeland 1993, pp. 17–20; Reisch 1991, p. 6; Huckfeldt 1990, p. 425).

Ivan Ermakoff - 9781789906851

6.2 Interdependence

Open contingency brings along a different class of empirical referents than the Cournot contingency. The focus is now on conjunctures defined in light of actors' own involvement in, and experience of, an action system. Validating claims pertaining to such conjunctures requires documenting their temporal boundaries. That is, we should be able to identify the moments in time when the defining characteristics of these conjunctures take shape as well as the moments when they recede to the background to the point of becoming inexistent. The following remarks tackle this issue for the three types distinguished earlier: action systems susceptible of a sudden conflagration, alternative evolutionary paths and collective irresolution. These remarks underscore the potential heuristic benefit that can be derived from combining morphological indicators (that is, indicators highlighting formal patterns) with an analysis of actors' situated experience.

6.2.1 Susceptibility to cascades

Regarding action systems presumed to be on the cusp of undergoing an outbreak of disruptive contention, the task of validation amounts to assessing the possibility of a large-scale cascade without the privilege of hindsight. Action systems may of course be more or less prone to sudden shifts in state as a result of their systemic properties, be these observable directly as topological characteristics (Albert et al. 2000; Watts 2002) or indirectly through phenomenal patterns, such as a size distribution of disturbances following a power law (Bak 1996, p. 27; Brunk 2001, pp. 430–3; Biggs 2005, pp. 1697–1703).[10] From this systemic and morphological standpoint, the possibility of deflagration is coterminous with the commencement of a global cascade (Watts 2002, p. 5768).

At issue is whether one can trace the temporal emergence of a conjuncture prone to deflagrations before a large-scale disruptive cascade starts. Three indicators a priori have relevance: (1) a shift downward of action threshold values, (2) an increase in the frequency of disturbances and (3) the emergence of shared beliefs geared to transgressive collective action. While each indicator can reveal growing tension within the action system, each also raises challenges of its own. For instance, data highlighting actors' conditional propensities for transgressive contention—their action thresholds—is usually very difficult to obtain especially in an authoritarian context (Kuran 1991, p. 124).

In the absence of reliable data documenting shifts in action thresholds, it may seem at first glance reasonable to try to gauge whether an action system is undergoing growing tensions by paying attention to the temporal evolution of disturbances (e.g. the pace of demonstrations in the Soviet Union starting in the summer 1987: Beissinger 2002, pp. 74–91). Still, an increase in disturbances does not necessarily mean that the action system has reached a point when a deflagration has become a possibility. Relying on contemporary ethnographic observations may prove very useful to make this assessment provided that these observations highlight median subjective dispositions within the collective under consideration. Trotsky (1932) points to observations of that kind when, about the "commotion" of the May 1917 demonstrations in

[10] It should be noted that the informational content of an indirect "signature" (Bak et al. 1988: p. 364) of system-level properties may be open to question. As Frigg (2003) points out, "the finding of a power law can, but need not be, an indication of an underlying self-organized critical structure" (p. 624).

Ivan Ermakoff - 9781789906851

Saint Petersburg, he remarks: "an attentive ear might have caught already among the ranks of the workers and soldiers impatient and even threatening notes" (p. 346).

6.2.2 Stochastic paths

Validating claims of evolutionary indeterminacy implies several empirical assessments. The first bears upon the existence and the modalities of one or several positive feedback mechanisms. The indeterminacy of outcomes rests on the self-amplifying effects of these mechanisms in an interactive environment in which the Cournot contingency, both in the singular and plural modes, is present (Arthur 1994, p. 27). Under these conditions, interactions produce "contingent outcomes, that is, outcomes in which random events are not averaged away" (Hill 1997, p. 190). Documenting positive feedback is therefore key. The second assessment focuses on the outset of the process for the purpose of gauging whether incidental factors or arbitrary conjunctions gave shape to the dynamic of adoption (e.g. Cowan 1990, pp. 559–64). The third assessment maps out the realm of the possible as historical actors viewed and experienced it. This mapping of possibilities identifies which options were available to actors at the outset of the process. The "vantage point" is ex ante (David 1985, p. 334).

6.2.3 Markers of uncertainty

Mutual uncertainty provides the empirical gradient of the third type of open-ended conjuncture—collective indecision. Not only are individuals uncertain about what they should think and do. The uncertainty displayed by those who face the same decisional challenge feed their own. This basic observation sets the agenda for the purpose of validating empirical claims. It invites us to identify the behavioral markers of mutual uncertainty. Do actors adopt a wait-and-see attitude? Do they oscillate as they revise expectations about the stance of their reference group? Are they unwilling to disclose action preferences in public or in interpersonal interactions? Do they betray an interest in aligning their stance with one another's (e.g. Ermakoff 2008, chapters 8 and 9)?

7. CONDITIONS AND IMPACTS

An empirical understanding of contingency heightens theoretical stakes by making it more difficult to take refuge in assertoric statements. This observation takes on particular significance when we investigate social change. The Cournot contingency raises the question of *how* happenstance and random variation affect collective dynamics. Open contingency invites us to research which properties make action systems open to contrasted collective outcomes. For the purpose of investigating social change, these two empirical foci—impact and conditions—complement one another. We cannot hope to analyze collective dynamics unless we theorize how impacts shape conditions and, conversely, in which respects conditions mediate impacts.

Providing a fully fledged treatment of the impact/condition nexus is beyond the scope of this chapter. The aim of the following considerations is only to show that an empirical and modal understanding of contingency has much to contribute to this perspective of research. I do so by considering three analytical foci: (1) single individual acts, (2) local interactions and (3) group experience. The focus on single individual acts highlights how relational configurations condition the impact of happenstance. The study of random interactions shows how their impact varies depending on whether they set the first stage of an evolutionary path, operate as

Ivan Ermakoff - 9781789906851

triggers of multiple cascades, or initiate a shift in equilibria. The analysis of group experience draws attention to the conditions leading to the emergence of collective irresolution and the bracketing of structural relations.

7.1 Incidental Agency

Accidents affecting individual agency can shape collective outcomes in different ways (Ermakoff 2015, pp. 78–80). The impact is *epistemic* when an individual action affects beliefs about the immediate future. For illustrative purpose, it is sufficient to mention the statement made on 9 November 1989 by a member of the East German Communist Party (SED) Politbüro about the possibility for East German citizens to cross the border with West Germany. In response to questions by journalists, this official, Guenther Schabowski, in charge of announcing the new travel regulations devised by the SED Central Committee, stated that these regulations allowed East German citizens to cross checkpoints along the Berlin Wall without restriction and with immediate effect. This statement, which was actually mistaken—Schabowski "had just returned to Berlin, and had not yet seen the directive or taken part in any discussion about it"—gave way to the widespread belief that the wall was now open (Kramer 2011, p. 843–4).

The impact of happenstance through individual agency is *network-based* as a result of the actor's positional characteristics in a structure of ties. Let us mention two configurations for illustrative purposes: (1) an individual actor who connects two separate groups is de facto the venue through which these two groups can communicate with one another. Happenings that undercut this individual's ability to communicate, mediate or bargain has collective consequences; (2) scale-free networks, that is, networks in which the degree distribution is a power law, undergo fragmentation when the most connected individuals are incapacitated (Albert et al. 2000, pp. 380–1).

Individual action, or the lack thereof, has a *pivotal* impact when it changes a "precarious" balance of forces between two camps opposing each other (Leff 1969, p. 52). For instance, one defection is enough to turn winning and losing positions upside down. Decisional procedures may be such that the acquisition of a pivotal position turns out to be an emergent feature of the process of stance taking. Consider the situation faced by the French Convention Assembly in the evening of 17 January 1793: the deputies had to decide whether death was to be the punishment meted out to the king, Louis XVI, for his acts of treason. The vote proceeded as a roll-call: one by one representatives publicly stated their stance. Death without reprieve won by only one vote (Ozouf 1989, pp. 102–3; Tackett 2015, pp. 240–1).

A *pyramidal* impact takes place in action systems in which actors have formally transferred control over their own actions to one individual or a set of individuals (Coleman 1990, p. 66). Hierarchical organizations are prone to pyramidal impacts when those to whom a right of control has been granted provide the organization with its impulsion and motion. Incidental changes affecting these actors' agency have repercussions on subordinates and their agents. This impact can be metaphorically described as pyramidal because it trickles down to the whole organization.

An individual action has a *sequential* impact when it elicits a process of behavioral alignment: actors mimic the reaction already displayed by others. The chain of events that led to the opening of the wall on 9 November 1989 illustrates this process. Earlier in the day, the announcement that East Germans could cross the border without restriction led thousands of East Berliners to gather at the Bornholmer crossing point. Since no instruction had been

Ivan Ermakoff - 9781789906851

conveyed to them, border guards administratively treated those requesting passage to the west as permanent emigrants and stamped their identification cards with this status, so much so that when these East Berliners came back in the evening to go home, entrance was initially denied to them. People shouted. The tension rose up. Unable to reach superiors who could tell them what to do, the local guards at the Bornhomer checkpoint decided to allow free passage. Soon afterward, guards at other crosspoints followed suit (Maier 1997, pp. 160–1). The wall had in effect become open.

7.2 Random Interactions

Interactions that are random in the sense that they are primarily coincidental affect social dynamics in various ways. The difficulty lies in figuring out a framework providing analytical leverage to tackle the roles they play. To this end, I propose to distinguish three classes of impacts depending on whether the focus is on the initial stage of a dynamic process, the triggering of cascades or the transition phase between two equilibria.

That incidental interactions at an early stage can be highly consequential is one of the conclusions yielded by strategic accounts of path dependency mentioned in Sections V and VI (Arthur 1994, p. 117; Hill 1997, pp. 198–9). We need not endorse an evolutionary argument in order to emphasize the conditional significance of random variations at the onset of a process. Studies exploring the empirical relevance of chaos theory applied to social processes have made the same point (Roth 1992 p. 209; Brown 1997, p. 128). So has experimental work centered on the signals delivered by early qualitative assessments (Salganik et al. 2006, p. 855).

Random interactions can also prove significant in triggering chain reactions (cascades) of various scales, which over time contribute to the emergence of a conjuncture prone to large-scale chain reactions. In this line of argument, triggering effects are not confined to a specific scale even though they are particularly noticeable when the cascade is large scale (Biggs 2005, pp. 1686–8). Finally, random deviations from interactional patterns can initiate shifts in equilibria. At times actors incidentally deviate from the decision rule that usually governs their behavioral choices when they interact with others (Young 1998, p. 19; DellaPorta et al. 2017, p. 11). As they cumulate, these deviations create the conditions for a shift in interactional patterns. A new equilibrium emerges. Its genealogy, however, is stochastic.

7.3 Critical Challenges

The focus on group experience points to the significance of decisional challenges, which actors experience as a *collective* process. Actors are caught off guard or taken aback because they lack either the analytical grid allowing them to make sense of the challenge, or the script to mount a planned response. Nonetheless, their uncertainty also betrays a shared awareness that their response—interpretive or behavioral—needs to be a collective one. They seek to achieve a common understanding of what is going on, and they aspire to a collective response. For this purpose, they are willing, if not eager, to make their interpretation and their behavioral response conditional on the stances displayed, or suggested, by those whom they view as referential. In these conditions, structural determinations are up in the air. Depending on which stance will emerge from the process, "structures" understood as enduring patterns, can be tossed up, altered or reinstated.

Ivan Ermakoff - 9781789906851

8. CONCLUSION

This chapter expounded a modal conception of contingency as encompassing different modes of indeterminacy, that is, different ways for happenings and conjunctures of being deprived of necessity regarding their cause (happening) or their outcome (conjuncture). A modal understanding had three defining characteristics that are of particular relevance for the analysis of contingency. First, it implies a realist approach to its referents. There would be no point in distinguishing different "ways of being" or "ways of being effectuated" if we could not assume that the phenomenal reality displaying these different modes has an existence independent of the act of conceptualizing them. Second, a modal understanding places the empirics of its cases at the center of the stage. Accordingly, this chapter took pain to relate conceptual claims to the empirics of specific cases. Third, in line with its concern for issues of demarcation and differentiation, a modal understanding is geared to the specification of conditional factors. All three features, as I hope to have made clear, are key to an empirical understanding of the roles played by contingency's various guises in social dynamics.

REFERENCES

Albert, R., H. Heong and A.L. Barabási (2000), "Error and Attack Tolerance of Complex Networks", *Nature*, **406**, 378–82.

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

Bak, Per (1996), *How Nature Works: The Science of Self-Organized Criticality*, New York: Springer.

Bak, Per, Chao Tang and Kurt Wiesenfeld (1988), "Self-Organized Criticality", *Physical Review*, A **38** (1), 364–76.

Ballinger, Clint (2013), "Contingence, déterminisme et 'just-so stories'", *Tracés*, **24** (1), 47–69.

Beissinger, Mark (2002), *Nationalist Mobilization and the Collapse of the Soviet State*, Cambridge: Cambridge University Press.

Biggs, Michael (2005), "Strikes as Forest Fires: Chicago and Paris in the Late Nineteenth Century", *American Journal of Sociology*, **110**, 1684–714.

Boudon, Raymond (1986), *Theories of Social Change. A Critical Appraisal*, Berkeley: University of California Press.

Bouveresse, Jacques (1993), *L'Homme probable: Robert Musil, le hasard, la moyenne et l'escargot de l'histoire*, Combas: Édition de l'éclat.

Brown, Thad A. (1997), "Measuring Chaos Using the Lyapunov Exponent", in Douglas L. Kiel and Euel Elliott (eds), *Chaos Theory in the Social Sciences: Foundations and Applications*, Ann Arbor: Michigan University Press, pp. 53–66.

Brucker, Gene A. (2001), "The Horseshoe Nail: Structure and Contingency in Medieval and Renaissance Italy", *Renaissance Quarterly*, **54** (1), 1–19.

Brunk, Gregory G. (2001), "Self-Organized Criticality: A New Theory of Political Behaviour and Some of Its Implications", *British Journal of Political Science*, **31** (2), 427–45.

Brunk, Gregory G. (2002), 'Why do Societies Collapse? A Theory Based on Self-Organized Criticality', *Journal of Theoretical Politics*, **14** (2), 195–230.

Bury, J.B. (1930), *Selected Essays*, Cambridge: Cambridge University Press.

Carr, Edward Hallett (1963), *What Is History?* New York: Alfred Knopf.

Chalmers, T.W. (1952), *Historic Researches. Chapters in the History of Physical and Chemical Discovery*, New York: Scribner.

Collins, Randall (1993), "Emotional Energy as the Common Denominator of Rational Action", *Rationality and Society*, **5** (2), 203–30.

Collins, Randall (2004), *Interaction Ritual Chains*. Princeton, NJ: Princeton University Press.

Coleman, James S. (1966), "Foundations for a Theory of Collective Decisions", *American Journal of Sociology*, **71** (6), 615–27.

Coleman, James S. (1990), *Foundations of Social Theory*, Cambridge, MA: Harvard University Press.

Cournot, Augustin (1851), *Essai sur les fondements de nos connaissances et sur les caractères de la critique philosophique*, (1975), Paris: Vrin.

Ivan Ermakoff - 9781789906851

Cowan, Robin (1990), "Nuclear Power Reactors: A Study in Technological Lock-in", *The Journal of Economic History*, **50** (3), 541–67.
Cowan, Robin and Philip Gunby (1996), "Sprayed to Death: Path Dependence, Lock-in and Pest Control Strategies", *The Economic Journal*, **106** (486), 521–42.
David, Paul A. (1985), "Clio and the Economics of QWERTY", *The American Economic Review*, **75** (2), 332–7.
DellaPosta, Daniel, Victor Nee and Sonja Opper (2017), "Endogenous Dynamics of Institutional Change", *Rationality and Society*, **29**, 5–28.
Ekeland, Ivar (1993), *The Broken Dice and Other Mathematical Tales of Chance*, Chicago: University of Chicago Press.
Elster, Jon (1998), "A Plea for Mechanisms", in Peter Hedström and Richard Swedberg (eds), *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge: Cambridge University Press: 45–73.
Ermakoff, Ivan (2001), "Strukturelle Zwänge und zufällige Geschehnisse" [Structural Constraints and Accidental Happenings], *Geschichte und Gesellschaft*, Sonderhefte 19: Struktur und Ereignis: 224–56.
Ermakoff, Ivan (2008), *Ruling Oneself Out. A Theory of Collective Abdications*, Durham, NC: Duke University Press.
Ermakoff, Ivan (2010), "Motives and Alignment", *Social Science History*, Special Section: Politics, Collective Uncertainty, and the Renunciation of Power, **34** (1), 97–109.
Ermakoff, Ivan (2015), "The Structure of Contingency", *American Journal of Sociology*, **121** (1), 64–125.
Ermakoff, Ivan (2017), "Shadow Plays: On Theory's Perennial Challenges", *Sociological Theory*, **35** (2), 128–37.
Ermakoff, Ivan and Marko Grdesic (2019), "Institutions and Demotions: Collective Leadership in Authoritarian Regimes", *Theory and Society*, **48** (4), 559–87.
Ertman, Thomas (1997), *Birth of the Leviathan. Building States and Regimes in Medieval and Early Modern Europe*, Cambridge: Cambridge University Press.
Fearon, James D. (1996), "Causes and Counterfactuals in Social Science: Exploring an Analogy between Cellular Automata and Historical Processes', in Philip E. Tetlock and Aaron Belkin (eds), *Counterfactual Thought Experiments in World Politics*, Princeton, NJ: Princeton University Press: 39–67.
Frigg, Roman (2003), "Self-Organised Criticality—What It Is and What It Isn't", *Studies in History and Philosophy of Science*, **34**, 613–32.
Gallie, William Byers (1964), *Philosophy and the Historical Understanding*, New York: Schocken.
Glasser, Otto (1959), *Wilhelm Conrad Röntgen und die Geschichte der Röntgenstrahlen*, Zweite Auflage, Berlin-Göttingen-Heidelberg: Springer.
Goldthorpe, John H. (2016), *Sociology as a Population Science*, Cambridge: Cambridge University Press.
Grafstein, Robert (1988), "A Realist Foundation for Essentially Contested Political Concepts", *Western Political Quarterly*, **41** (1), 9–28.
Granovetter, Mark (1978), "Threshold Models of Collective Behavior", *American Journal of Sociology*, **83** (6), 1420–43.
Griffin, Larry J. (1993), "Narrative, Event-Structure Analysis, and Causal Interpretation in Historical Sociology", *American Journal of Sociology*, **98** (5), 1094–133.
Griffin, Larry J. and Charles C. Ragin (1994), "Some Observations on Formal Methods of Qualitative Analysis", *Sociological Methods and Research*, **23** (1), 4–21.
Griffin, Larry J. and Robert R. Korstad (1995), "Class as Race and Gender: Making and Breaking a Labor Union in the Jim Crow South", *Social Science History*, **19** (4), 425–54.
Hall, John R. (1999), *Cultures of Inquiry: From Epistemology to Discourse in Sociohistorical Research*, Cambridge: Cambridge University Press.
Hawthorne, Geoffrey (1991), *Plausible Worlds. Possibility and Understanding in History and the Social Sciences*, Cambridge: Cambridge University Press.
Haydu, Jeffrey (2010), "Reversals of Fortune: Path Dependence, Problem Solving, and Temporal Cases', *Theory and Society*, **39** (1), 25–48.
Hedström, Peter (2005), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge: Cambridge University Press.
Hill, Greg (1997), "History, Necessity and Rational Choice Theory", *Rationality and Society*, **9** (2), 189–213.
Huckfeldt, Robert (1990), "Structure, Indeterminacy and Chaos: A Case for Sociological Law", *Journal of Theoretical Politics*, **2** (4), 413–33.
Isaac, Larry W., Debra A. Street and Stan J. Knapp (1994), "Analyzing Historical Contingency with Formal Methods. The Case of the 'Relief Explosion' and 1968", *Sociological Methods and Research*, **23** (1), 114–41.
Isaac, Larry W. and Larry J. Griffin (1989), "Ahistoricism in Time-Series Analyses of Historical Process: Critique, Redirection, and Illustrations from U.S. Labor History", *American Sociological Review*, **54** (6), 873–90.
Kramer, Mark (2011), "The Demise of the Soviet Bloc", *The Journal of Modern History*, **83** (4), 788–854.
Kreps, David M. (1990), *Game Theory and Economic Modeling*, Oxford, UK: Clarendon Press.
Kuhn, Thomas S. (1996), *The Structure of Scientific Revolutions*, Chicago and London: University of Chicago Press.

Ivan Ermakoff - 9781789906851

Kuran, Timur (1991), "The East European Revolution of 1989: Is It Surprising That We Were Surprised?', *The American Economic Review*, **81** (2), 121–5.
Lebow, Richard Ned (2000), "Contingency, Catalysts, and International System Change", *Political Science Quarterly*, **115** (4), 591–616.
Leff, Gordon (1969), *History and Social Theory*, London: Merlin.
Macy, Michael and Milena Tsvetkova (2015), "The Signal Importance of Noise", *Sociological Methods & Research*, **44** (2), 306–28.
Mahoney, James (2000), 'Path Dependence in Historical Sociology', *Theory and Society*, **29**, 507–48.
Maier, Charles S. (1997), *Dissolution: The Crisis of Communism and the End of East Germany*, Princeton, NJ: Princeton University Press.
Manzo, Gianluca (2014), "Data, Generative Models, and Mechanisms: More on the Principles of Analytical Sociology", in Gianluca Manzo (ed.), *Analytical Sociology: Actions and Networks*, Chichester: Wiley: 4–52.
Merton, Robert K. (1968), *Social Theory and Social Structure*, New York: The Free Press.
Ozouf, Mona (1989), "King's Trial", in François Furet and Mona Ozouf (eds), Arthur Goldhammer (trans.), *A Critical Dictionary of the French Revolution*, Cambridge, MA: Belknap Press: 95–105.
Pascal, Blaise (1670), *Pensées*, reprinted in M. Le Guern (ed.) (2004), Paris: Galllimard.
Pearson, Karl (1904), "On the Theory of Contingency and Its Relation to Association and Normal Correlation", Drapers' Company Research Memoirs, Biometrics Series I, Department of Applied Mathematics, University College, University of London, London: Dulau.
Ragin, Charles C. (1987), *The Comparative Method. Moving Beyond Qualitative and Quantitative Strategies*, Berkeley and Los Angeles, CA: University of California Press.
Reisch, George A. (1991), "Chaos, History, and Narrative", *History and Theory*, **30** (1), 1–20.
Richards, Diana (1997), "From Individuals to Groups: The Aggregation of Votes and Chaotic Dynamics', in Douglas L. Kiel and Euel Elliott (eds), *Chaos Theory in the Social Sciences: Foundations and Applications*, Michigan, IL: Michigan University Press: 89–116.
Roth, Randolph (1992), "Is History a Process? Nonlinearity, Revitalization Theory, and the Central Metaphor of Social Science History", *Social Science History*, **16** (2), 197–243.
Ruelle, David (1991), *Chance and Chaos*, Princeton, NJ: Princeton University Press.
Sahlins, Marshall (1985), *Islands of History*, Chicago, IL: University of Chicago Press.
Salganik, Matthew J., Peter Sheridan Dodds and Duncan J. Watts (2006), "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market", *Science*, **311** (5,762), 854–6.
Schedler, Andreas (2007), "Mapping Contingency", in Ian Shapiro and Sonu Bedi (eds), *Political Contingency: Studying the Unexpected, the Accidental, and the Unforeseen*, New York: New York University Press: 54–78.
Sewell, William H. Jr. (2005), *Logics of History. Social Theory and Social Transformation*, Chicago, IL: University of Chicago Press.
Shapiro, Ian and Sonu Bedi (2007), "Introduction", in Ian Shapiro and Sonu Bedi (eds), *Political Contingency: Studying the Unexpected, the Accidental, and the Unforeseen*, New York: New York University Press: 1–18.
Sohrabi, Nader (2011), *Revolution and Constitutionalism in the Ottoman Empire and Iran*, Cambridge: Cambridge University Press.
Somers, Margaret (1998), "We're No Angels: Realism, Rational Choice, and Relationality in Social Science", *American Journal of Sociology*, **104** (3), 722–84.
Spinoza, Baruch (1966), trans. by Charles Appuhn, *Oeuvres IV*, Paris: Flammarion.
Steinmetz, George (1998), "Critical Realism and Historical Sociology. A Review Article", *Comparative Studies in Society and History*, **40** (1), 170–86.
Straus, Scott (2006), *The Order of Genocide. Race, Power and War in Rwanda*, Ithaca, NY: Cornell University Press.
Tackett, Timothy (2015), *The Coming of the Terror in the French Revolution*, Cambridge, MA: Belknap Press.
Trotsky, Leon (1932), Max Eastman (ed. and trans.) (1999), *History of the Russian Revolution*, New York: Pathfinder.
Watts, Duncan J. (2002), "A Simple Model of Global Cascades on Random Networks", *Proceedings of the National Academy of Sciences*, **99**, 5766–71.
Wolfram, Stephen (1984), "Universality and Complexity in Cellular Automata", in J. Doyne Farmer, Tommaso Toffoli and Stephen Wolfram (eds), *Cellular Automata: Proceedings of an Interdisciplinary Workshop*, Amsterdam: North Holland Physics: 1–35.

Ivan Ermakoff - 9781789906851

15. Cumulative advantage

Freda B. Lynn and Hannah W. Espy

INTRODUCTION

The idea of cumulative (dis)advantage resonates deeply in the social sciences, in part because it is easily recognizable in so many dimensions of social life. With respect to a collection of actors or objects—e.g. a class of workers, a set of products, a cohort of students—we know that their pathways over the life or career course will likely diverge with respect to both monetary and non-monetary rewards, such as social status or health in the case of individuals. Not only do we expect the rich to get richer, we know from the myriad competitive systems in which we participate, that winners, given enough time, will likely "take all" (Frank and Cook 1995).

Not surprisingly, studies of cumulative advantage (henceforth CA) are authored by scholars from an array of disciplines and address a wide variety of phenomena.[1] Topics include the economics of stardom (Adler 1985; Rosen 1981), rise of status hierarchies (Gould 2002; Manzo and Baldassarri 2015), racial barriers to labor market success (Blau and Duncan 1967; DiPrete 1981), dynamics of scientific careers (Azoulay and Lynn 2020; Bol et al. 2014; Merton 1968), consequences of curricular tracking in schools (Carbonaro 2005; Lucas 2001), downward spirals in health (Ross and Wu 1996; Willson et al. 2007), and various network-based externalities (see DiMaggio and Garip 2012).

This chapter does not offer a comprehensive review of the findings from this body of work but, rather, in the spirit of analytical sociology, organizes it around two key issues: group-level patterns and generative mechanisms. With regard to group-level patterns, CA refers to divergent trajectories for a group of actors with respect to some outcome Y, wherein some actors become more advantaged relative to others over time. But the pattern of divergence—a property of a group and not a trajectory—can differ across groups. That CA can manifest as different bundles of pathways is an issue long recognized (see Allison et al. 1982) but remains under-appreciated (see Bask and Bask 2015). An elaboration on pattern may spur the identification of more instances of CA than are understood currently. Thinking about patterns also paves the way for developing group-level questions about CA. For example, why do groups differ in the configuration of trajectories they experience?

Second, this chapter offers a framework for understanding the core suite of mechanisms capable of fueling divergence. Each mechanism represents "a set of entities and activities likely to trigger a sequence of events (a process) likely to bring about the outcome," which in this case refers to some form of CA (Manzo 2014, p. 7). While it can be relatively straightforward to document the presence of CA with respect to a given outcome, the force(s) that

[1] For a comprehensive overview, see DiPrete and Eirich (2006). See Zuckerman (1998) for an introduction to the theory of cumulative advantage and disadvantage over the life course given small initial differences; Adler (2006) for a model of how CA arises even in the absence of initial differences; DiMaggio and Garip (2012) for a review of how network effects interact with individual characteristics to exacerbate or ameliorate CA; and Sauder et al. (2012) and Sauder (2018) for discussions of system-level approaches to CA.

Freda B. Lynn and Hannah W. Espy - 9781789906851

generate it are not always readily observable, and hence difficult to tease apart. Thus, our goal is to distill from the current literature a "generative model" (Hedström 2005; Manzo 2014) of CA, by which we mean the mechanisms that could produce divergence even if acting alone. For each mechanism, we first illustrate the basic idea and then discuss the theoretical and empirical work advocating its importance for CA. Ultimately, we address each mechanism in isolation to reiterate that the CA observed "on the surface" is likely sustained by a complex host of interacting forces underneath. We close the chapter by proposing next steps for research on cumulative (dis)advantage.

1. PATTERNS OF DIVERGENCE

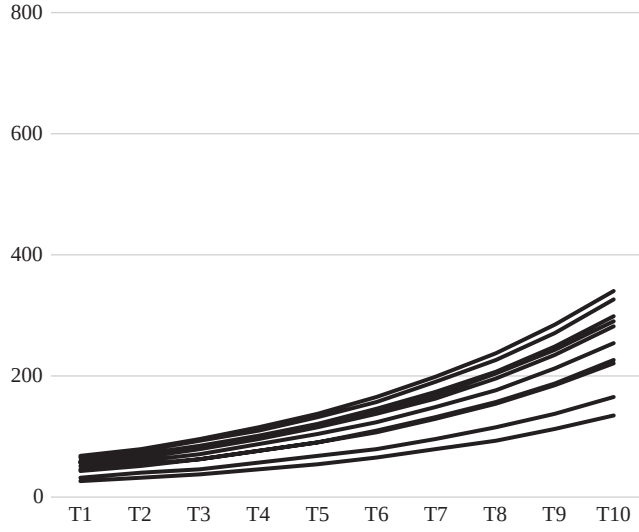
In generic form, CA describes a situation in which one actor becomes more advantaged relative to another over time. If rewards accrue faster to actor A compared to actor B, then as time progresses, A's advantage over B accumulates. Thus, CA is a process pertaining to a collection of trajectories (i.e. two or more actors) and not a single actor, and its basic elements are time and divergence. Of course, as Allison et al. (1982) remind us, "the simplicity of this formulation belies the diversity and complexity of the underlying processes" (p. 615).

In this section, we focus on how these two key elements—time and divergence—tell the story of a collection of actors. To simplify, we consider trajectories with respect to a single resource or reward, Y. These individual trajectories aggregate to establish a pattern of divergence (or convergence), and these patterns can differ with respect to key features. Once again, these features are group-level properties, and thus our discussion of patterns is about how groups (and not individuals) differ in their experience with CA. Summarized by Bask and Bask (2015, p. 6), there is, on the one hand, *intra-individual change* (a micro-level phenomenon) and on the other, *inter-individual divergence* (or convergence) in pathways (a macro-level phenomenon). Emphasized below are three important ways macro-level systems of CA vary.

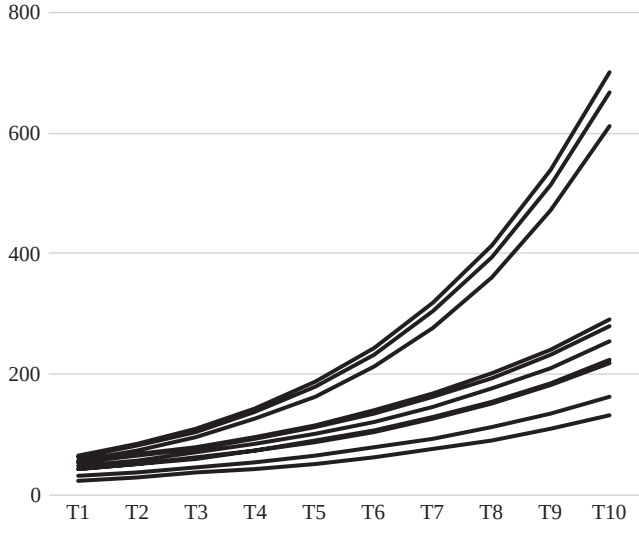
Constant versus Differential Rates of Returns. The classic discussion of CA focuses on growth rates of some reward Y. As discussed in DiPrete and Eirich (2006), a variety of formal models could capture the CA process depending on one's conceptualization of the mechanisms producing CA. A key consideration in this modeling process is whether the rate of growth in Y is constant or varies across actors. Although many simply assume that CA means growing inequality, Allison et al. (1982) clarify that CA will only generate an increase in inequality if growth rates differ across actors. If growth rates are constant, absolute differences in reward values will grow, but they will remain proportional to initial differences. Figure 15.1a illustrates a 10-actor system characterized by proportional growth, whereas 1b illustrates a system in which inequality increases over time. In Figure 15.1a, the coefficient of variation (see Allison et al. 1982) is equivalent at time 1 and 10. In Figure 15.1b, the most advantaged actors get rich faster (30% interest rate) compared to those with lesser advantage (20% interest rate), and the coefficient of variation at time 10 is 2.3 times greater than in time 1.

Incremental Change versus Discrete Shifts. Perhaps because variable interest rates are such intuitive examples of CA, one might easily assume CA to be a process inherently characterized by incremental divergence. Indeed, Figures 15.1a–b are apt summaries of CA in many empirical settings, such as the divergence in cumulative grant funding (Bol et al. 2018) and

Freda B. Lynn and Hannah W. Espy - 9781789906851



a. Fixed Growth Rate



b. Variable Growth Rate

Notes: Trajectories are generated based on a path-dependent CA process (see DiPrete and Eirich 2006, equation 7): $Y_{it} = (1 + \gamma)Y_{i,t-1}$. $Y_{i,t=1}$ is the same in 15.1a and 15.1b and was drawn randomly from a uniform distribution on the interval (20, 80). In Figure 15.1a, $\gamma = .20$ and is constant across all actors. In Figure 15.1b, $\gamma = .30$ for the top three actors and $\gamma = .20$ for the bottom seven. The coefficient of variation in 15.1a at $t = 1$ and $t = 10$ is .27. In Figure 15.1b, the coefficient of variation in $t = 10$ is .61.

Figure 15.1 Incremental divergence

Freda B. Lynn and Hannah W. Espy - 9781789906851

divergence in cumulative citation counts (Azoulay and Lynn 2020) within cohorts of scientists. But not all CA processes unfold this way, and some patterns of CA may be better characterized with respect to discrete shifts in divergence.

For example, Figures 15.2a–b offer two conceptualizations of how a collection of pathways can diverge by way of critical junctures. Figure 15.2a, for example, illustrates emergent inequality via the top few shifting tracks, leaving the rest behind. Specifically, the top three in time 4 are catapulted to a new strata of success and then the most advantaged actor enjoys a second discrete boost at time 8, creating greater inequality not only for the entire group, but within the elite. This type of CA process, which hinges on only some actors being elevated to an elite class, aligns conceptually with Merton's (1968) discussion of the 41st chair phenomenon and his use of the phrase "ratchet effects," as well as Accominotti's (2018) work on consecration as a population-level phenomenon, in which the glorification of a select few allows a collective to project hierarchical orderliness.

Figure 15.2b illustrates the same idea but with respect to the presence of downward stepwise trajectories, i.e. suffering at the bottom rather than glorification at the top. Figure 15.2b resonates with research on workers experiencing "scarring effects" in the labor market (Ellwood 1982, see also Blank et al. 2004), as well as the "mark" of stigmatized positions (Pager 2003).

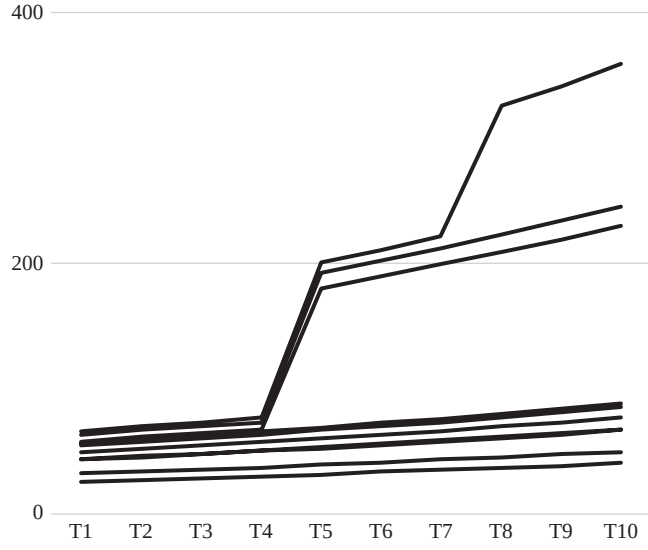
A recent empirical study of the gender gap in grant funding among life scientists (Lerchenmueller and Sorenson 2018) makes a strong case for examining CA with regard to track shifts. In the context of the amount of time it takes scientists to receive a coveted R01 grant, which is largely understood as the transition point to becoming an independent researcher, they document that "a large share of [the gender] gap emerges in a relatively short period of time" around the R01 transition, and that in contrast to the imagery of "women dripping out of the STEM career pipe every centimeter along the way, they appear to pour out at one of the critical junctures" (p. 1015).

Exaggeration with and without Decoupling. The last macro-level feature we highlight is decoupling. In Figures 15.1a–b and 15.2a–b, initial differences are magnified over time; in Figure 15.3, initial differences in Y are magnified over time and, simultaneously, the rank order of actors shifts over time. Put differently, the group's hierarchy at time 1 and time 10 are decoupled (see Lynn et al. 2009).[2] The issue of decoupling falls outside traditional CA, but is an important feature to consider moving forward because it highlights how the CA process is not just about actors locking in advantage over time, but more broadly about actors being relegated to qualitatively different pathways.

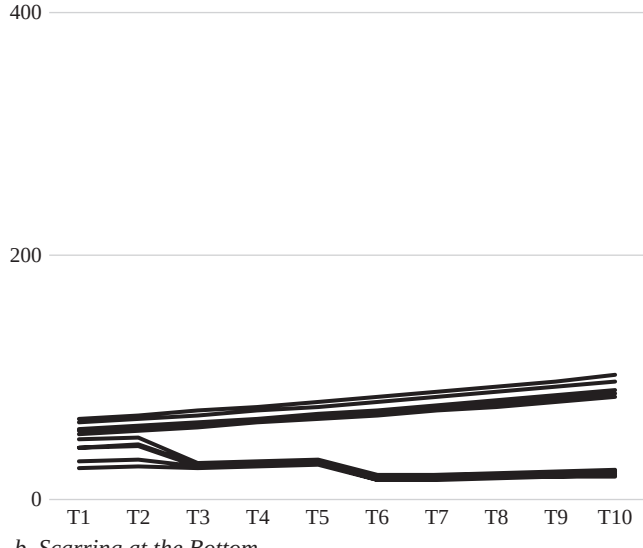
In Figure 15.3, for example, actors land in positions that are not explicable based on their initial positions. Actor A, for example, pulls far ahead of the pack despite starting in the middle. In contrast, actor C is unable to convert initial advantage into downstream advantage and worse, actor D loses advantage over the long term. For example, if time 10 is an election and time 1 the start of the race, then candidates B and D "peaked too early"; despite a fast rate of positive growth during periods 1–5, a counter mechanism not only stopped their momentum but led to a reversal of fortune. In other words, in some situations, the mechanisms

[2] The model shown in Figure 15.3 is one in which grow rates vary across actors and within actors over time, which generates non-monotonic trajectories.

Freda B. Lynn and Hannah W. Espy - 9781789906851



a. Elevation at the Top



b. Scarring at the Bottom

Notes: The distribution of $Y_{i,\,t=1}$ is the same as Figure 15.1a–b. The bottom seven trajectories in 15.2a and the top five in 15.2b are given by $Y_{it} = (1.05)Y_{i,\,t-1}$. At $t = 10$, the coefficient of variation is .83 in Figure 15.2a and .66 in Figure 15.2b.

Figure 15.2 Divergence via tracking

Freda B. Lynn and Hannah W. Espy - 9781789906851

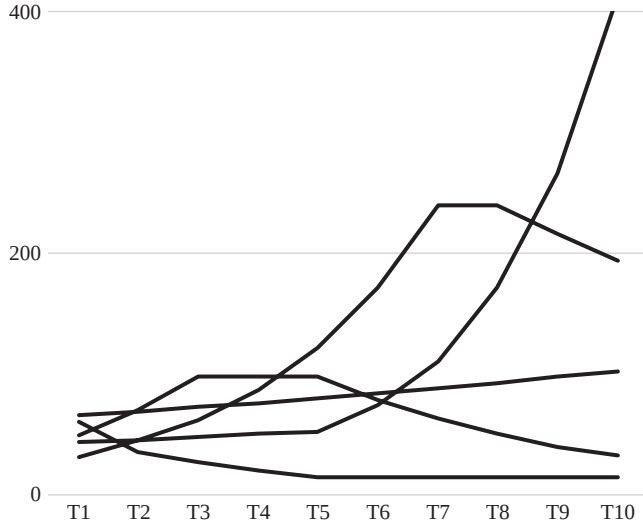


Figure 15.3 Divergence and decoupling

that produce CA can be turned off or muted (see van de Rijt 2019 for a discussion of self-correcting dynamics). From this perspective, the study of CA could be redefined as the study of (1) heterogeneity in trajectories and (2) reversible and irreversible pathways.

2. MECHANISMS: EXPLAINING DIVERGENCE AND DECOUPLING

We turn now to explanations: What generates divergence and decoupling? Next, we highlight five fundamental mechanisms that we consider primary sources of fuel for CA. Again, in most empirical contexts, it is highly unreasonable to assign blame to a single mechanism. But isolating mechanisms as a theoretical exercise helps underscore the complexity underneath. To be clear, most theorists' accounts of CA involve multiple mechanisms. As surveys of the field have revealed (e.g. Adler 2006; DiPrete and Eirich 2006), theorists can be identified by how they define and (selectively) package mechanisms into a CA story.

2.1 Initial Endowments

Consider two runners. One runs a mile in four minutes and the other in eight. This difference in pace will obviously create divergence in distance covered in a single race, as well as divergence in long-term success in the sport more generally. Difference in pace could be attributed to superior physical and mental endowments, such as with respect to muscular strength, rate of maximal oxygen uptake (VO_2 max), mental stamina, and self-confidence.

Variation in endowments are widely discussed in theories of CA, although typically in combination with other mechanisms. Rosen's (1981) model of the superstar phenomenon, for example, is predicated on underlying variation in talent, which is defined

Freda B. Lynn and Hannah W. Espy - 9781789906851

as an actor's "ability in an activity" (p. 852). DiMaggio and Garip (2012) refer to initial endowments as the advantages individuals can access (at time 0), such as financial and cultural capital (p. 95). Gould (2002) refers to endowments as exogenous or intrinsic quality (Q), which he defines as whatever attributes "a social context makes salient as a determinant of social desirability" (p. 1153), such as an individual's performance, productivity, or contribution to group goals. Gould clarifies that intrinsic quality need not correspond to "anything of real intrinsic value" and instead only represents a "ranking on which actors can agree without reference to one another's judgments" (p. 1172). Interestingly, in lay accounts of "getting ahead" (e.g. Espy 2017; Hochschild 1995; McNamee 2004; Reynolds and Xian 2014) the notion of quality or merit is the central driver of why some surge past others, a theme we return to in the final section.

For some scholars, the point of discussing endowments is to clarify that divergence in trajectories can occur even if actors exhibit little to no such variation at the start (e.g. Adler 1985, 2006). This is precisely the point of many experimental studies whose purpose is to fix supply-side differences at zero in order to isolate the role of demand-side mechanisms (Correll et al. 2007; Delecourt and Ng 2019; Moss-Racusin et al. 2012). In the case of Salganik and colleagues' experimental music market (2006; 2008), they use a control condition to show that differences in the "intrinsic appeal" of songs cannot fully explain inequality in the downstream consumption of those songs. Adler (1985, 2006), countering Rosen (1981), argues that superstars can emerge in markets even if producers all possess the same talent.

In the context of educational stratification, Manzo (2013) makes the related but distinct point that variation in "ability" alone cannot plausibly explain the degree of class-based inequality we observe empirically with respect to educational attainment. Rational choice models suggest that students make educational transition decisions (e.g. "Should I persist or drop out?") by considering their academic ability, their probability of success given their ability, and the costs and benefits of pursuing more education in lieu of other options. From this perspective, ability (an initial endowment) is the key piece of information upon which each transition decision is based. In turn, variability in initial endowments should be the most important piece of the puzzle when trying to explain downstream variability in educational attainment. Yet Manzo (2013) shows with agent-based simulations that—even if student ability varies tremendously and in tandem with students' socio-economic background—it is implausible that ability differences alone will amplify to the degree of class-based educational inequality seen in true populations, unless (1) students' educational choices at each transition are influenced by their peers' decisions and (2) friendships tend to be homophilous with regard to ability. That is, it is only in the presence of network independencies that variation in initial endowment is "dynamically deepened and widened by means of a cumulative process" (2013, p. 73).

It bears reminding that endowments are not inherently exogenous and instead likely to be the product of CA processes themselves. That we treat endowments as "initial" is simply an artifact of research design. For example, the purpose of Lynn et al.'s (2009) model, which builds from Gould (2002), is to explain divergence in status *after time 0* among actors who differ at time 0 with respect to quality (Q); hence the origins of Q itself are left conveniently outside the scope of analysis.

Returning to the example above, the runner who is physically and mentally superior has more "talent" at time 0, but where does that talent come from? At time 0, physical form, for instance, is partly genetic gift but also a function of being born into a family with the financial

Freda B. Lynn and Hannah W. Espy - 9781789906851

resources to provide nutrition and the opportunity to nurture the development of muscular strength and coordination (see Fischer et al. 1996 on the endogeneity of ability). In addition, physical form can be nurtured by a family with the cultural savvy needed to navigate athletic training programs, identify the "hidden curriculum" of the sport, and build networks with coaches and trainers (see Calarco 2011, 2014; Lareau 2002, 2011). Physical form can also be shaped by an actor's network of neighbors, friends, and extended family—layers of social connections that can establish norms that promote healthy habits and protect from danger (see DiMaggio and Garip 2012; Jencks and Mayer 1990).[3]

Thus, even though it may be unfeasible to explain precisely the origins of endowments, there is value in conceptualizing initial endowments as all the resources actors have at their disposal at the start of the time period under examination. This includes traditional forms of capital (human, economic, cultural, and social), as well as any other deployable resource that is contextually relevant, such as space (see Doering and Liu 2019), bodily capital (Monk 2015), and organizational capital (Ellersgaard et al. 2019). Rates of returns can be contingent on any one of these resources.[4] Moreover, researchers should be mindful of their sampling time frame when studying CA processes, since time frame determines what counts as exogenous endowment. As suggested by Figures 15.2a–b, an observer would reach different conclusions about how much initial level of Y matters for CA depending on the slice of the timeline analyzed. Findings from one timeline segment may not generalize well to the "total" CA process.

2.2 System-Imposed Constraints

Institutionalized processes and system-wide constraints can induce divergent trajectories even if actors are seeded with the same initial endowments. Imagine two runners of equal talent at time 0. Nevertheless, an elite training program admits only one because only one meets the program's formal eligibility requirements. Over time, the runner accepted by the program pulls ahead of the one excluded. The eligibility criteria in this example constitute an institutionalized procedure that instigates divergence.

Institutionalized procedures include overt, legal mandates (e.g. 20th century Jim Crow laws and gender- and religion-based admissions restrictions in higher education in the United States), as well as hidden rules, such as the algorithms used to calculate insurance risk (Fan and Wang 2017; Joly et al. 2014). These institutionalized devices, often referred to in sociology as structural constraints, can be used to systematically favor some actors at the expense of others, thereby fueling CA. In a similar vein, system-imposed constraints can refer to the kind of exogenous, physical and physiological constraints on all actors (e.g. 24 hours in day, cognitive load constraints), but also the cultural constraints of widespread, consensual beliefs (e.g. Correll 2004), and market- and technological-based constraints, such as the generalized

[3] DiMaggio and Garip (2012) zoom in on how network effects can aggravate inequality in initial endowments because local networks affect the likelihood that an actor adopts a practice that promotes getting ahead. Under certain conditions, local network environments (e.g. first- and second-degree connections) can shape divergence in endowments to the extent that an advantage-conferring practice, such as eating healthy foods or staying in school, is more likely to be adopted in some clusters of actors, but not others. These adoption patterns thus concentrate advantage within only select pockets of the population.

[4] See DiPrete and Eirich (2006) equations 11–16 for status-resource interaction models.

Freda B. Lynn and Hannah W. Espy - 9781789906851

consumer's need for coordination (Adler 1985) or the fact that technology allowed art to become reproducible (Rosen 1981).

Discriminatory redlining policies in the insurance industry are one prominent example of how algorithms—the origins of which are often opaque—can make it easier for the advantaged to retain and build advantage. As summarized by Ito (2019):

> … insurance companies evolved and developed sophisticated-seeming calculations to show that their discrimination was "fair." Women should pay more for annuities because statistically they lived longer, and blacks should pay more for damage insurance when they lived in communities where crime and riots were likely to occur. While overt racism and bigotry still exist …, in insurance it has been integrated into and hidden from the public behind mathematics and statistics that are so difficult for nonexperts to understand that fighting back becomes nearly impossible.

Put differently, even if two actors are equal on "merit," overt and covert system-wide rules can set them on different pathways by how they values non-merit factors.

Similarly, examples abound of systems designed to give top performers an edge over the rest in competitive markets, which could likely result in a system akin to Figure 15.1b or 15.2a. Consider, for example, seeding in professional sports tournaments. A grand slam event such as Wimbledon relies on a formula that favors a certain type of top-performing player.[5] Ultimately, the seeding procedure ensures that currently lower-performing players have a steeper hill to climb compared to higher-performing players when trying to win the tournament (Della Croce et al. 2020; see also Scarf and Yusof 2011).

The very same principle is at play with respect to Google Scholar's search algorithm and the structuring of publication visibility. Efforts to reverse engineer Google's algorithm confirm that the most heavily weighted factor in terms of organizing search results is citation counts (Beel and Gipp 2009); specifically, articles that have been previously cited by many are most likely to appear at the top of the search results. This simple sorting mechanism creates a visibility and accessibility advantage based on popularity. Evans (2008) argues that electronic search, while efficient from the perspective of an individual scientist searching for publications, leads to "narrowing" in terms of the field of science. Online search "puts researchers in touch with prevailing opinion, but this may accelerate consensus and narrow the range of findings and ideas built upon" (p. 395). In sum, electronic search likely fuels winner-take-all dynamics in the market for attention in science.

Another way in which structural processes govern CA pertains to the phenomenon of consecration (Accominotti 2018) and, in particular, prize-giving and other institutionalized processes that do the work of legitimizing the rise of "the chosen." The number of competitors distinguished with a prize, for example, is an arbitrary limit imposed by the system and a socially constructed tool to create a tangible boundary separating winners and losers (Accominotti 2018; Allen and Parsons 2006). Examples of limit setting include the Olympic Games deciding in 1904 to award medals only to the top three finishers, Hall of Fame institutions' decisions to artificially cap the number of inductees per year, and the Nobel Foundation's decision to give one award in just six prize categories per year.

---

[5] http://www.tennisabstract.com/blog/2017/06/27/measuring-the-impact-of-wimbledons-seeding-formula/

Freda B. Lynn and Hannah W. Espy - 9781789906851

Merton (1968) famously brought this issue to light in his discussion of the French Academy and its practice of granting membership to only 40 "immortals" at a time, thereby relegating many brilliant minds to the unfortunate position of 41st chair. In this case, the "number of places available at the summit of recognition" (Merton 1968, p. 57) was set in the 1600s and has since been maintained by the French government. In sum, selective prize-giving appoints elite status to some but not others despite highly similar performance records. To the extent that winning confers legitimacy and creates opportunity denied to losers, the manufactured "scarcity of prizes" (Merton 1968) can fuel divergence independent of initial endowment.

2.3 Audience Valuation

If system-level constraints are imposed "top down," there are also ways in which CA is shaped from the "bottom up" through emergent processes (Gould 2002). Consider two runners of equal talent at time 0. Fans fall in love with one but not the other, arrive in droves at races, and draw sponsorship attention toward their beloved. Over time, the favored runner, validated and flush with sponsor support, pulls ahead of their competitor over the career course.

The favoritism in this example is a type of non-institutionalized, demand-side evaluation mechanism that can induce and sustain CA. Actors, ideas, and objects are routinely evaluated by others—peers, supervisors, and third parties—and not the system per se. Those favorably evaluated can pull ahead of those seen in a less favorable light, even if they are otherwise equivalent. From this perspective, stars are born not because of what they do, but when members of the audience elevate them to stardom.

Overall, audience members are in a position to (1) allocate resources, attention, and praise; (2) grant allowances; and (3) verify identity and validate reputation (Gould 2002; Merton 1968; Podolny 1993; Zuckerman 1999).[6] Different lines of sociological research demonstrate how audiences give and withhold this support, and how these evaluation processes fuel CA. Note that in contrast to the way in which actors are dealt with via system rules, audience evaluation is a local, emergent process, defined by some members of the audience (specific others) evaluating some actors some of the time in a particular way. Variation in perception means that one actor could fare well in one local environment but less so in another. With system-imposed constraints, however, actors with a certain configuration of characteristics will all be treated the same by the system until the rules are rewritten.[7]

Members of the audience, for example, can fuel divergence through the allocation of tangible resources.[8] When program officers of a funding agency award a grant to one scientist

---

[6] Audience effects are distinct from network effects. If, in this example, the favored runner is granted more exposure to elite training partners, then training with tougher competitors would constitute a type of network effect while the favoritism from coaches is rooted in audience evaluation. Again, network effects refer to how those connected to a focal actor, directly or indirectly, shape the trajectory of the focal actor because of the neighborhood (literal or metaphorical) they create. Audience effects, in contrast, pertain to how those in demand-side positions evaluate supply-side actors, which in bounded groups can be the same actors serving as both assessor and the assessed (Gould 2002).

[7] Only in the extreme, when nearly all members of the audience judge an actor with certain characteristics similarly in nearly all situations, would audience evaluation serve as a de facto system rule.

[8] Note that audience attention refers to the *discretionary* decision-making that drives who receives awards while system constraints on prize-giving pertain to matters such as the number of awards given, the nature of the award, and the formal criteria or algorithms used to determine award eligibility.

Freda B. Lynn and Hannah W. Espy - 9781789906851

and not another, the resources (time, money, lab equipment) associated with the grant could generate a future difference in the number of publications, even if the two scientists had nearly identical careers up until the time the award as given (e.g. Bol et al. 2018; Jacob and Lefgren 2011; Merton 1968). When managers allocate higher salary raises to non-minorities even when they have the same performance reviews as minorities (Castilla 2008), those monetary rewards can be used to maintain a higher level of mental and physical health that could lead to future differences in job performance.

The audience also allocates attention and praise—they back some but not others—and they do so in ways that can lead to divergence in tangible rewards downstream. Consumers, for example, create growing inequality in market systems when their decisions to "buy" some products and not others are influenced by the opinions of previous consumers (Salganik et al. 2006). Lynn et al. (2016a) find that products of mediocre intrinsic appeal are rated more likeable by consumers when presented as more versus less popular with respect to previous adopters.[9] This same halo effect is at play when members of hiring committees view sought-after candidates as smarter only because they are sought-after by others (Lynn et al. 2016b).

Ultimately, these valuation processes can create allowances for some actors that enable them to get ahead faster than others. For example, if consumers and other market actors tolerate higher-status organizations charging more for the same (Podolny 1993, 1994, 2005), this can lead to CA in profit. In the context of publishing, journal editors and reviewers contribute to gender-based CA when they hold (consciously or not) women to higher standards compared to men in the journal peer review process (Card et al. 2020; Hengel 2017), a practice that would ultimately pave the way for men to publish more. Likewise, to the extent that faculty still praise male applicants as more competent than otherwise equivalent female applicants for a job in a laboratory (Moss-Racusin et al. 2012), faculty allow men to feel more "at home" in laboratory jobs compared to women, providing a sense of security that better positions them for future productivity.

Audiences also play an important role in CA insofar as they have the power to verify an actor's identity. When an audience cannot make sense of the categories to which an actor or product belongs, research suggests that those actors and products will trade at a discount in the eyes of the audience (Zuckerman 1999; see also Hannan 2010, Zuckerman 2016).[10] In the labor market, for example, employers prefer candidates who have some diversity in their work histories, but disapprove of those who have skipped around too much between "really unrelated jobs" (Leung 2014, p. 143). Theories suggest that those who cannot be easily categorized are at risk of being perceived by the audience as illegitimate (Zuckerman 1999), inauthentic (Kovács et al. 2014), and lacking in quality overall (Hsu 2006; Hsu et al. 2009; Leung and Sharkey 2014; Pólos et al. 2002).

Finally, audience members can induce CA with respect to social status by how they form direct attachments to, and hence socially validate, some but not others. Actors can be incorporated, as well as shunned, through the energy that others are willing to invest in them (Gould 2002; Manzo and Baldassarri 2015). Miller et al. (2021) argue that part of the struggle that

[9] Research demonstrates that socially influenced decision-making can not only exacerbate outcome inequality (Manzo and Baldassarri 2015) but, in certain circumstances, such as when self-correcting dynamics are prevented, lead to the elevation of suboptimal choices (van de Rijt 2019).

[10] In Zuckerman's (2016) revised account of optimal distinctiveness, an audience evaluates candidates through a two-stage valuation process, in which candidates must first conform to categorical partitions (to be assigned a comparison group) but then distinguish themselves within that comparison group.

Freda B. Lynn and Hannah W. Espy - 9781789906851

historically black colleges and universities (HBCUs) face to get ahead in the field of higher education is that historically white institutions (HWIs) fail to recognize HBCUs as legitimate peer institutions in the field of education. Being excluded from the inner circle leaves HBCUs without an important form of peer support (e.g. knowledge sharing) that could help HBCUs gain footing (see also Wooten 2015).

In summary, decentralized acts of valuation on the part of the actors themselves can induce divergence in a coveted outcome Y within a set of actors. That is, even if candidates demonstrate equal merit and are treated equally by the system an audience can find ways to value them differently by their willingness (conscious or not) to accept and validate some but not others.

2.4 Self-Selection

Imagine two runners of equal talent who are similarly embraced by the audience. Both are offered spots in an elite training program but only one decides to enroll. The one who opted in, over time, outperforms the one who declined the opportunity, and hence self-selection could be viewed as the critical reason for divergence in their trajectories over time.

Actors' decisions to opt in versus out is a distinct, supply-side source of fuel for CA. To be sure, highlighting the role of self-selection mechanisms is not an exercise in blaming the victim (Correll 2004), but rather an attempt to show how these decision-making processes should be recognized as a unique lever in the CA process. In some cases, self-selection may be tightly linked to other mechanisms, such as when widespread cultural beliefs, a type of implicit system constraint, shape self-selection decisions. But culture does not pre-determine agency. Crucial steps in the CA story are the actors' *own* decisions to deploy resources *in response to* system constraints and audience valuation.

Correll (2001, 2004), for example, convincingly argues that cultural beliefs about gender and STEM lead girls to make downwardly biased self-assessments about their math competence, which in turn shapes decision to persist in STEM careers: "boys do not pursue mathematical activities at a higher rate than girls do because they are better at mathematics. They do so, at least partially, because they *think* they are better" (2001, p. 1724). This is an example of cultural constraints converting (in a probabilistic sense) into individual-level preferences. As Correll rightly points out, "attempts to counter the effects of gender beliefs on gender segregation and inequality in the labor force … will require looking beyond how stereotypes are used by gatekeepers, such as teachers and employers" (an audience effect) and instead must also incorporate "how gender beliefs affect males' and females' perceptions of their own abilities at crucial decision-making junctures" (self-selection) (pp. 1725–6).

Another major theme pertaining to self-selection is the observation that "winners" tend to invest more effort into winning future contests. Carbonaro's (2005) work on tracking in schools argues that curricular tracking shapes academic outcomes not only because tracking poses structural constraints on learning, but also because tracking affects human agency, which in this case refers to the effort students allocate to learning. Effort here is not a proxy for values or beliefs but rather pertains to energy expenditure to meet academic requirements. Carbonaro finds that students in more advanced tracks exert more effort than students in lower tracks, and this is the case even after accounting for the fact that students who exerted more effort in the past are more likely to be tracked higher. In other words, it is the experience of participating in a more competitive track that extracts greater investment in learning. Additional social psychological research is needed, however, to know if this effect is driven simply by students

Freda B. Lynn and Hannah W. Espy - 9781789906851

wanting to "up their game" to meet the challenge of a more competitive program or if elite tracks are structured to be greedier than their participants. In either case, winning leads actors to allocate more effort to staying ahead.[11]

Bol et al. (2018) notice a similar dynamic in their examination of an academic funding system, in which they find a strong CA pattern that cannot be reduced to variation in researcher quality. Using a regression discontinuity approach, they compare grant winners whose review scores are just above the funding threshold versus losers just below the cut-off. Results show that initial winners proceed in the ensuing eight-year period to receive more than twice the total funding received by initial losers. Importantly, their research design allows them to document that while part of the observed CA is driven by demand-side evaluation (i.e. funders view past winners more favorably than losers), self-selection also contributes to divergence. Grant winners are more likely to keep applying for more grants whereas losers tend to shy away from future competition. The authors refer to this as a "participation" mechanism.[12] It is not that award recipients and non-recipients are endowed unequally with respect to self-motivation, but rather that non-recipients decide to allocate their motivation elsewhere (they decide not to participate) upon losing in an early stage in the tournament.

2.5 Chance, Randomness, and Luck

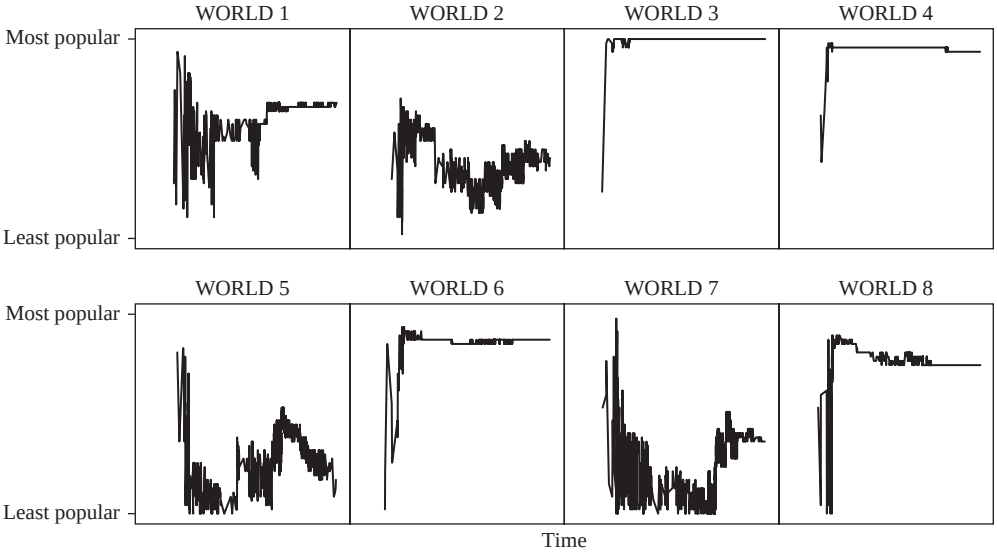
Two runners with equal talent and equal access to training opportunities are valued equally by the audience. Both opt into an elite training program and are similarly invested in their success. The day before training starts, the two cross the street and one is hit by a speeding car and breaks her leg in three places. The healthy runner, over time, pulls ahead of the injured runner, who is sidelined for twelve months and never fully regains strength in one knee.

On one hand, the role that luck, chance, and randomness play in CA has not gone unacknowledged in theories of CA. Merton and Barber (2011) and Merton (1945) discuss the role of "serendipity" in scientific discovery, and hence, the careers of scientists. Jencks et al. (1972) famously claimed that a large portion of income inequality could be attributable to luck: "chance acquaintances who steer you to one line of work rather than another, the range of jobs that happen to be available in a particular community when you are job hunting, the amount of overtime work in your particular plant, whether bad weather destroys your strawberry crop, whether the new superhighway has an exit near your restaurant, and a hundred other unpredictable accidents" (p. 227). Arthur (1989) argues the importance of random historical events in explaining the rise of one technology over another. In accordance with these ideas, DiPrete

[11] This aligns with Merton's (1968) observation of how peer expectations can motivate self in the case of science: "… social pressures do not often permit those who have climbed the rugged mountains of scientific achievement to remain content. It is not necessarily … that their own Faustian aspirations are ever escalating that keeps eminent scientists at work. More and more is expected of them, and this creates its own measure of motivation and stress. Less often than might be imagined is there repose at the top in science" (p. 57).

[12] In a twist on this argument, Lynn et al. (2019) argue that one of the reasons why there is a muted gender gap in citations per publication yet a major gender gap in number of publications is because self-selection mechanisms are virtually eliminated in the context of citations. Scientists cannot opt out of receiving citations once a manuscript is published insofar as the audience primarily gets to decide whether to cite or ignore. In contrast, there are many self-selection junctures (in addition to demand-side hurdles) when it comes to publishing an article. An author must decide, for example, if an idea is worth developing into a manuscript, if a working paper is ready for feedback, if a draft is ready to submit for peer review, if reviewers' comments are (in)surmountable, and if a revision is ready for re-submission.

Freda B. Lynn and Hannah W. Espy - 9781789906851



Notes: The y-axis corresponds to market rank and the x-axis refers to real time in each world. "Lockdown" is rated as above average in the control condition (not shown). Regardless of rating, however, most of the 48 songs exhibit a great deal of variation in trajectory and destination positions across their eight lives. The only major exception pertains to the song with the very highest rating and the one with the very lowest rating (see Salganik and Watts 2008). For example, for the very highest rated song, "She Said" by Parker Theory, seven of its eight lives in Experiment 2 are very similar, where it rises to the top immediately and locks in its advantage for the duration of the world. But, for the 46 remaining songs in between the two extremes, their eight lives exhibit variation similar to that of "Lockdown."

Source: See Salganik and colleagues (2006, 2008) and https://opr.princeton.edu/archive/cm/.

Figure 15.4 Eight lives of one song, "Lockdown" by 52metro, in Musiclab Experiment 2

and Eirich (2006) present models of CA in which "exogenous chance events" are folded into a path-dependent CA model (see equation 7, pp. 276–7).

On the other hand, perhaps with the exception of Arthur (1989), CA theories tend to avoid treating luck and randomness as keynote mechanisms, such as how Rosen (1981) treats variation in talent and Gould (2002) social influence. Yet, many CA stories could be rewritten with luck in the starring role. Consider, for example, the Musiclab experiment (Salganik et al. 2006; Salganik and Watts 2008), a highly influential study upheld for demonstrating the role of social influence in the creation of inequality. One could easily argue, however, that the results shine an equal spotlight on the importance of luck and randomness in understanding CA.

To illustrate, Figure 15.4 focuses on one of 48 songs in the experimental market, and displays its "life" course in each of the eight different social influence conditions to which participants were assigned.[13] The x-axis represents time and the y-axis represents market rank,

[13] In each social influence condition, respondents were shown song titles, band names, and the number of times each song had been downloaded previously by respondents in their world. In each world, all songs started with zero downloads and then accumulated downloads in real time. Hence, we can observe eight socially influenced "lives" for each song in the experiment (see Lynn et al. 2016).

Freda B. Lynn and Hannah W. Espy - 9781789906851

where position 1 refers to the most downloaded song and position 48 corresponds to the least downloaded. Figure 15.4 shows that the pathway of "Lockdown" varies tremendously across worlds, all of which are subject to social influence. In other words, while it is certainly true that social influence (and intrinsic appeal) plays a role in shaping a song's life (Salganik et al. 2006; Salganik and Watts 2008), it would also be fair and true to conclude that random chance decides which life a song gets to live.[14]

The Musiclab experiment beautifully illustrates this point given its creation of multiple counterfactual lives per song. In most empirical settings, we observe only one trajectory per actor, and typically only a short segment at that. This perhaps is one explanation for why the role of luck and randomness is often undersold in discussions of CA. In the absence of counter-factual observations, the importance of luck and chance is difficult to "see" let alone measure. Consequently, we gravitate toward proving and disproving locally specific CA explanations to make sense of the one reality we do observe, rather than entertain the causes of different potential realities (see Laland et al. 2011). On this point, computational simulation is a particularly promising method for, among other benefits (Manzo 2014), helping researchers appreciate the role that randomness plays in the dynamics of complex systems.[15]

More generally, luck and randomness are also undersold because such explanations seem to remain unsatisfying to both researchers and the public. Sauder (2020) offers an extensive discussion of how luck is viewed as "antithetical to a sociological approach to understanding the world" (p. 197).[16] He describes, for example, how Jencks et al.'s (1972) reliance on luck as an explanatory factor was met with controversy (Sauder 2020, p. 196), and how, more generally, sociologists tend to "scrub luck from the story of inequality" (p. 200). In contrast, theorists in other fields, such as evolutionary biology, embrace luck and randomness as critical components of theory building (see Lieberson and Lynn 2002).

Similarly, the importance of luck is relatively downplayed across cultures in terms of public opinion. Figure 15.5, for example, shows nationally representative PEW data (n = 48,643 from 44 countries) on the perceived importance assigned to "hard work" versus "luck" with regard to getting ahead (see Espy 2017). In Figure 15.5, the x-axis plots the mean importance assigned to hard work and the y-axis the mean importance of luck, both of which were measured on a 10-point scale (10 = most important). A country would thus fall on the diagonal line if it assigned luck and hard work the same level of importance. Figure 15.5 shows that (1) hard work is never rated below a 7.5, whereas luck sometimes is, and that (2) luck is rated by most countries as less important or far less important than hard work.[17]

---

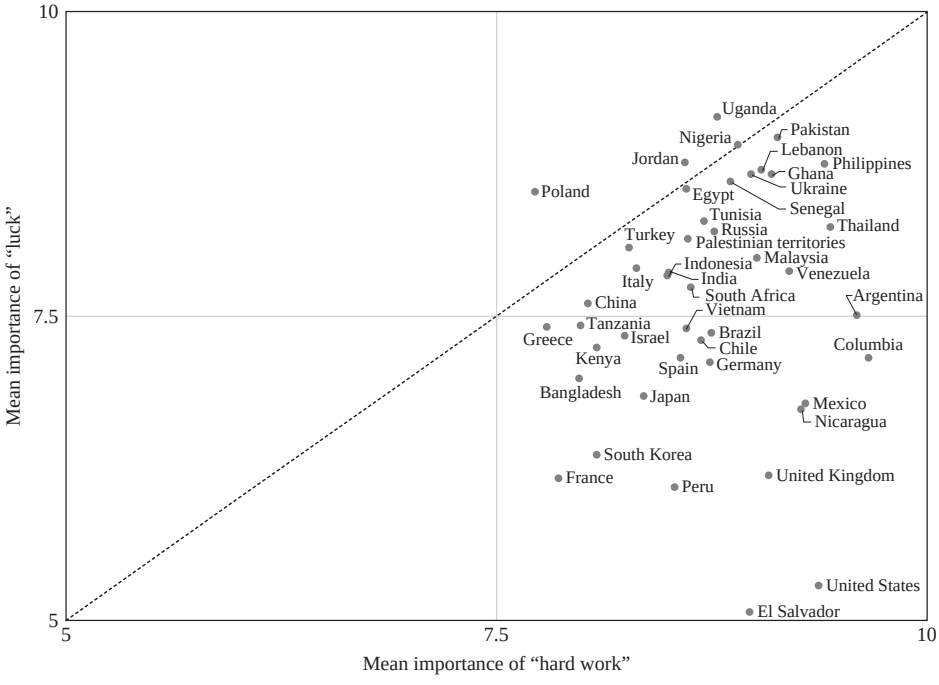
[14] Specifically, the chance element in this situation corresponds to the random assignment of participants to worlds. These participants arrive in random order and their (socially influenced) choices generate a trajectory.

[15] See also Macy and Tsvetkova (2015) and their argument regarding the theoretical importance of noise for understanding complex systems. In this review, they argue that science is traditionally viewed as the business of peeling away noise to find latent, non-random patterns (p. 307). Noise is cast as hindrance and associated with "inefficiency, unpredictability, and diversity" (p. 308). To prove otherwise, Macy and Tsvetkova (2015) review studies demonstrating how randomness in complex systems can actually generate outcomes that are, counter-intuitively, more predictable and less diverse.

[16] There is, however, a fast-growing body of literature emerging around luck and chance and its role in social and professional systems (see Denrell 2005; Denrell and Liu 2012; Frank 2016; Liu and de Rond 2016; Sinatra et al. 2016).

[17] See also Denrell and Liu (2012, pp. 9, 335) for different kinds of evidence of luck being downplayed.

Freda B. Lynn and Hannah W. Espy - 9781789906851



Notes: Axes correspond to a country's mean ranking of the importance of hard work (x-axis) and luck (y-axis) for "getting ahead" in life. A country falls on the diagonal line if luck and work are rated as equally important. The majority of the 44 countries surveyed view hard work as more important than luck.

Source: Pew Research Center, *Spring 2014 Global Attitudes survey*.

Figure 15.5 Public perceptions of "getting ahead"

To be clear, it is not just luck that is downplayed in favor of meritocratic explanations; individual effort and "earned" achievement, such as educational credentials, are consistently ranked as more important than all other factors (Espy 2017). A meritocracy implies a system in which those endowed with merit or talent (or those disciplined enough to invest in developing merit) rise to the top *for only those* reasons. Overall, the combination of strong beliefs in hard work paired with relatively weaker beliefs in all alternative explanations is evidence of an over-reliance on meritocratic explanations at the expense of others.

While this pattern holds generally across cultures, there are groups for which it is notably strong, such as the United States compared to other countries, and conservatives compared to liberals within the US. There is also evidence that meritocratic explanations are used to explain variation in "starting points" (for example, hard work and not luck is the reason someone is better educated or has more social capital at T0). Within the United States specifically, belief in the importance of hard work for success is strongly correlated with the importance of education and "knowing the right people" (Espy 2017). The correlation between these three factors suggests that an individual's educational achievements and network position are attributable to individual-level merit factors (for example, hard work in school, intelligence, or superior networking strategies). Once again, other mechanisms such as system constraints and luck are left out of the story.

Freda B. Lynn and Hannah W. Espy - 9781789906851

One reason for this particularly strong affinity toward meritocratic explanations within the US is likely due to the public's long-standing embrace of individualism and "personal fitness" (Pease et al. 1970) vis-à-vis the American Dream ideology (Cullen 2003; Hochschild 1995). These ideologies promote successes achieved through merit as morally superior, while framing achievement through other means as un-American (Cullen 2003; Hochschild 1995).[18] As a result, Americans often struggle to disentangle what *does* matter and what *should* matter when thinking about success, attaching negative moral judgements to successes framed in terms of luck, personal identity (for example, race, class, or gender), or bribery.

In sum, the dominant explanation of how people get ahead across cultures, especially in the US, is highly aligned with a meritocratic logic. However, as suggested by this chapter and elsewhere (see Frank 2016; McNamee 2004), the concept of a meritocracy is a highly anemic account of CA. A more general issue is that meritocratic thinking favors the myopic interpretation of winners and losers as having earned their rewards (see Merton 1968 on the "injustice" of these attribution errors).[19] Overall, the disconnect between public perceptions of getting ahead and social science research on how CA actually operates is baffling.

3. MOVING FORWARD

Research on cumulative advantage has grown tremendously since Merton's (1968) discussion of the "Matthew effect" in science. Insights on how CA emerges and operates can be traced to a strikingly diverse pool of scholars, including business specialists studying employee compensation, sociologists of education examining curricular tracking schools, health scholars studying the cascading effects of poverty on mortality, and economists modeling the emergence of superstars in the entertainment industry. Cataloging patterns and fundamental mechanisms helps us manage the unwieldiness that comes with many "cooks in the kitchen," but more importantly helps reiterate how the combination of elements highlighted, isolated, and ignored yields distinctive accounts.

Moving forward, researchers would benefit from a conversation about CA and proximate versus ultimate causes à la Mayr (1961) (see also Laland et al. 2011). Such a framework could help us better understand how these mechanisms work together simultaneously, interactively, or consecutively. Currently, empirical research tends to be designed with the goal of verifying—yay or nay—if one particular mechanism is "on" at a given moment with respect to resource Y. But perhaps some mechanisms are at play within certain segments of the timeline but muted elsewhere. Similarly, perhaps certain mechanisms govern divergence with some outcomes but not others. Explaining total CA then would require changing lenses partway through, and/or zooming out and focusing on how different mechanisms work for different resources. Note that upon zooming out we are better positioned to discuss how different

---

[18] See Pease et al. (1970) on how even the sociological study of stratification in the US was slow to move away from individualistic (and optimistic) explanations of mobility.

[19] In a similar vein, Adler (2006) describes the double frustration of an unpopular artist: "Because the number of artists who can be popular at any one time is limited, not all talented artists can be successful. However, the frustration of an unsuccessful artist does not end with not having an audience. She must also suffer consumers' judgment that she deserves her fate. It is easy to conclude that an artist who is not as popular is not as good. Thus, the hierarchy of success manufactures a hierarchy of pseudo-talent in its own image, tending to reinforce the spurious perception of a talent differential" (p. 897).

Freda B. Lynn and Hannah W. Espy - 9781789906851

mechanisms interact or mutually sustain one another. We see important glimpses of this kind of thinking already, such as with respect to how cultural constraints (system constraints) convert into preferences (self-selection) (Correll 2001) and the way that audience biases become encoded via hidden algorithms (Noble 2018).

Zooming out even more, the field has much to gain by devoting more attention to the question of how CA *systems* differ and why it matters for the actors embedded within. Currently, we are flush with empirical and theoretical work on the mechanisms that can generate divergence itself, but we lack coherent theory as to why *patterns* of divergence differ across groups (see, however, Sauder 2006, Lynn et al. 2009, Sauder et al. 2012, Manzo and Baldassarri 2015, and Accominotti and Tadmon 2020 for discussions of group-level variation). For example, why does one group come to resemble Figure 15.3 while another Figure 15.2a? Again, current research is highly focused on explaining why a particular resource is characterized by cumulative (dis)advantage over a specific time period within a particular group. Asking if mechanism X generates CA in Y does not inform us as to why mechanism X is activated in some groups but not others.

In addition, because groups are not typically the unit of analysis (again, see above for exceptions), we know very little about the impact of living within different kinds of CA systems. There is country-level research, for example, on the effects of income inequality (see Wilkinson and Pickett 2009), but their focus is on living with greater or lesser levels of inequality.[20] This is distinct from asking how actors fare under different systems of CA. For example, does an organization or profession in which career trajectories resemble Figure 15.3 generate more or less innovative capacity compared to an organization or profession in which trajectories resemble Figure 15.2a? Likewise, while CA predicts a growing gap between winners and losers no matter the pattern, are actors better or worse off in terms of, say, aggregate social welfare (e.g. frustration, happiness) in one type of system versus another? How aware are actors that the systems in which they are embedded can be characterized by their mobility patterns? We are already well-versed in explaining how individual actors move from origin to destination positions and accumulate (dis)advantage over others, but less is known about what gives rise to an entire configuration of mobility patterns. The frontier is understanding why groups differ in the kind of CA they experience.

REFERENCES

Accominotti, Fabien (2018), "Consecration as a Population-Level Phenomenon", *American Behavioral Scientist*. https://doi.org/10.1177/0002764218800144.

Accominotti, Fabien and Daniel Tadmon (2020), "How the Reification of Merit Breeds Inequality: Theory and Experimental Evidence". Working Paper, London School of Economics.

Adler, Moshe (1985), "Stardom and Talent", *The American Economic Review*, **75** (1): 208–12.

Adler, Moshe (2006), "Stardom and Talent", in V.A. Ginsburg and D. Throsby (eds), *Handbook of the Economics of Art and Culture*, Vol. 1, Amsterdam: Elsevier, pp. 895–906.

Allen, Michael Patrick and Nicholas L. Parsons (2006), "The Institutionalization of Fame: Achievement, Recognition, and Cultural Consecration in Baseball", *American Sociological Review*, **71** (5): 808–25.

Allison, Paul D., J. Scott Long and Tad K. Krauze (1982), "Cumulative Advantage and Inequality in Science", *American Sociological Review*, **47** (5): 615–25.

Arthur, W. Brian (1989), "Competing Technologies, Increasing Returns, and Lock-In by Historical Events", *The Economic Journal*, **99** (394): 116–31.

[20] See also Heiserman and Simpson (2017) for an experimental approach to investigating how actors perceive merit in the context of high, medium, and low income inequality.

Azoulay, Pierre and Freda B. Lynn (2020), "Self-Citation, Cumulative Advantage, and Gender Inequality in Science", *Sociological Science*, **7**: 152–86.
Bask, Miia and Mikael Bask (2015), "Cumulative (Dis)Advantage and the Matthew Effect in Life-Course Analysis", *PLoS ONE*, **10** (11).
Beel, Jöran and Bela Gipp (2009), "Google Scholar's Ranking Algorithm: An Introductory Overview", *Procedures of the International Society for Scientometrics and Informetrics*, 230–41.
Blank, Rebecca M., Marilyn Dabady and Constance F. Citro (eds) (2004), *Measuring Racial Discrimination*, Washington, DC: National Academies Press.
Blau, Peter M. and Otis Dudley Duncan (1967), *The American Occupational Structure*, New York: Wiley.
Bol, Thijs, Jacqueline Witschge, Herman G. Van de Werfhorst and Jaap Dronkers (2014), "Curricular Tracking and Central Examinations: Counterbalancing the Impact of Social Background on Student Achievement in 36 Countries", *Social Forces*, **92** (4): 1545–72.
Bol, Thijs, Mathijs de Vaan and Arnout van de Rijt (2018), "The Matthew Effect in Science Funding", *Proceedings of the National Academy of Sciences of the United States of America*, **115** (19): 4887–90.
Calarco, Jessica McCrory (2011), "'I Need Help!' Social Class and Children's Help-Seeking in Elementary School", *American Sociological Review*, **76** (6): 862–82.
Calarco, Jessica McCrory (2014), "The Inconsistent Curriculum Cultural Tool Kits and Student Interpretations of Ambiguous Expectations", *Social Psychology Quarterly*, **77** (2): 185–209.
Carbonaro, William (2005), "Tracking, Students' Effort, and Academic Achievement", *Sociology of Education*, **78** (1): 27–49.
Card, David, Stefano DellaVigna, Patricia Funk and Nagore Iriberri (2020), "Are Referees and Editors in Economics Gender Neutral?", *The Quarterly Journal of Economics*, **135** (1): 269–327.
Castilla, Emilio J. (2008), "Gender, Race, and Meritocracy in Organizational Careers", *American Journal of Sociology*, **113** (6): 1479–526.
Correll, Shelley J. (2001), "Gender and the Career Choice Process: The Role of Biased Self-Assessments", *American Journal of Sociology*, **106** (6): 1691–730.
Correll, Shelley J. (2004), "Constraints into Preferences: Gender, Status, and Emerging Career Aspirations", *American Sociological Review*, **69** (1): 93–113.
Correll, Shelley J., Stephen Benard and In Paik (2007), "Getting a Job: Is There a Motherhood Penalty?", *American Journal of Sociology*, **112** (5): 1297–339.
Cullen, Jim (2003), *The American Dream: A Short History of an Idea That Shaped a Nation*, New York: Oxford University Press.
Delecourt, Solène and Odyssia Ng (2019), "Do Buyers Discriminate against Female-Owned Businesses? Two Field Experiments", Working Paper, Stanford University.
Della Croce, Federico, Gabriele Dragotto and Rosario Scatamacchia (2020), "On Fairness and Diversification in WTA and ATP Tennis Tournaments Generation", *Annals of Operations Research*.
Denrell, Jerker (2005), "Should We Be Impressed with High Performance?", *Journal of Management Inquiry*, **14** (3): 292–8.
Denrell, Jerker and Chengwei Liu (2012), "Top Performers Are Not the Most Impressive When Extreme Performance Indicates Unreliability", *Proceedings of the National Academy of Sciences*, **109** (24): 9331–6.
DiMaggio, Paul and Filiz Garip (2012), "Network Effects and Social Inequality", *Annual Review of Sociology*, **38** (1): 93–118.
DiPrete, Thomas A. (1981), "Unemployment Over the Life Cycle: Racial Differences and the Effect of Changing Economic Conditions", *American Journal of Sociology*, **87** (2): 286–307.
DiPrete, Thomas A. and Gregory M. Eirich (2006), "Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments", *Annual Review of Sociology*, **32** (1): 271–97.
Doering, Laura B. and Christopher C. Liu (2019), "From the Ground Up: Gender, Space, and Self-Employment in a Colombian Housing Project", *Sociology of Development*, **5** (2): 198–224.
Ellersgaard, Christoph Houman, Jacob Aagaard Lunding, Lasse Folke Henriksen and Anton Grau Larsen (2019), "Pathways to the Power Elite: The Organizational Landscape of Elite Careers", *The Sociological Review*, **67** (5): 1170–92.
Ellwood, David (1982), "Teenage Unemployment: Permanent Scars or Temporary Blemishes?", in *The Youth Labor Market Problem: Its Nature, Causes, and Consequences*, Chicago: University of Chicago Press, pp. 349–90.
Espy, Hannah (2017), "Makin' It: Mapping Logics of Success in American Culture", Unpublished Master's Thesis, Department of Sociology, University of Iowa, Iowa City, IA.
Evans, James A. (2008), "Electronic Publication and the Narrowing of Science and Scholarship", *Science*, **321** (5887): 395–9.
Fan, Chiang Ku and Wei-Yuan Wang (2017), "A Comparison of Underwriting Decision Making Between Telematics-Enabled UBI and Traditional Auto Insurance", *Advances in Management & Applied Economics*, **7** (1): 17–30.
Fischer, Claude S., Martin Sanchez Jankowski, Michael Hout, Samuel R. Lucas, Ann Swidler and Kim Voss (1996), *Inequality by Design: Cracking the Bell Curve Myth*, Princeton, NJ: Princeton University Press.

Freda B. Lynn and Hannah W. Espy - 9781789906851

Frank, Robert H. (2016), *Success and Luck: Good Fortune and the Myth of Meritocracy*, Princeton, NJ: Princeton: University Press.
Frank, Robert H. and Philip J. Cook (1995), *The Winner-Take-All Society Hardcover*, New York, The Free Press.
Gould, Stephen Jay (2002), *The Structure of Evolutionary Theory*, Cambridge, MA: Harvard University Press.
Hannan, Michael T. (2010), "Partiality of Memberships in Categories and Audiences", *Annual Review of Sociology*, **36** (1): 159–81.
Hedström, Peter (2005), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge, UK: Cambridge University Press.
Heiserman, Nicholas and Brent Simpson (2017), "Higher Inequality Increases the Gap in the Perceived Merit of the Rich and Poor", *Social Psychology Quarterly*, **80** (3): 243–53. doi: 10.1177/0190272517711919.
Hengel, Erin (2017), "Publishing While Female: Are Women Held to Higher Standards? Evidence from Peer Review", University of Cambridge, Cambridge UK, Cambridge Working Paper Economics Series #1753.
Hochschild, Jennifer L. (1995), *Facing Up to the American Dream*, Princeton, NJ: Princeton University Press.
Hsu, Greta (2006) "Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production", *Administrative Science Quarterly*, **51** (3): 420–50.
Hsu, Greta, Michael T. Hannan and Özgecan Koçak (2009), "Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests", *American Sociological Review*, **74** (1): 150–69.
Ito, Joi (2019), "Supposedly 'Fair' Algorithms Can Perpetuate Discrimination", *MIT Media Lab*. Retrieved 17 March 2020, www.media.mit.edu/articles/supposedly-fair-algorithms-can-perpetuate-discrimination/
Jacob, Brian A. and Lars Lefgren (2011), "The Impact of NIH Postdoctoral Training Grants on Scientific Productivity", *Research Policy*, **40** (6): 864–74.
Jencks, Christopher, Marshall Smith, Henry Acland, Mary Jo Bane, David Cohen, Herbert Gintis, Barbara Heyns, and Stephan Michelson (1972), *Inequality; a Reassessment of the Effect of Family and Schooling in America*, New York: Basic Books.
Jencks, Christopher and Susan E. Mayer (1990), "The Social Consequences of Growing up in a Poor Neighborhood", in *Inner-City Poverty in the United States*, Washington, DC: National Academy Press, pp. 111–86.
Joly, Yann, et al. (2014), "Life Insurance: Genomic Stratification and Risk Classification", *European Journal of Human Genetics*, **22** (5): 575–9.
Kovács, Balázs, Glenn R. Carroll and David W. Lehman (2014), "Authenticity and Consumer Value Ratings: Empirical Tests from the Restaurant Domain", *Organization Science*, **25** (2): 458–78.
Laland, Kevin N., Kim Sterelny, John Odling-Smee, William Hoppitt and Tobias Uller (2011), "Cause and Effect in Biology Revisited: Is Mayr's Proximate-Ultimate Dichotomy Still Useful?", *Science*, **334** (6062): 1512–16.
Lareau, Annette (2002), "Invisible Inequality: Social Class and Childrearing in Black Families and White Families", *American Sociological Review*, **67** (5): 747–76.
Lareau, Annette (2011), *Unequal Childhoods: Class, Race, and Family Life, Second Edition with an Update a Decade Later*, Berkeley: University of California Press.
Lerchenmueller, Marc J. and Olav Sorenson (2018), "The Gender Gap in Early Career Transitions in the Life Sciences", *Research Policy*, **47** (6): 1007–17.
Leung, Ming D. (2014), "Dilettante or Renaissance Person? How the Order of Job Experiences Affects Hiring in an External Labor Market", *American Sociological Review*, **79** (1): 136–58.
Leung, Ming D. and Amanda J. Sharkey (2014), "Out of Sight, Out of Mind? Evidence of Perceptual Factors in the Multiple-Category Discount", *Organization Science*, **25** (1): 171–84.
Lieberson, Stanley and Freda B. Lynn (2002), "Barking up the Wrong Branch: Scientific Alternatives to the Current Model of Sociological Science", *Annual Review of Sociology*, **28** (1): 1–19.
Liu, Chengwei and Mark de Rond (2016), "Good Night, and Good Luck: Perspectives on Luck in Management Scholarship", *Academy of Management Annals*, **10** (1): 409–51.
Lucas, Samuel R. (2001), "Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects", *American Journal of Sociology*, **106** (6): 1642–90.
Lynn, Freda B., Mary C. Noonan, Michael Sauder and Matthew A. Andersson (2019), "A Rare Case of Gender Parity in Academia", *Social Forces*, **98** (2): 518–47.
Lynn, Freda B., Joel M. Podolny and Lin Tao (2009), "A Sociological (De)Construction of the Relationship between Status and Quality", *American Journal of Sociology*, **115** (3): 755–804.
Lynn, Freda, Brent Simpson, Mark Walker and Colin Peterson (2016b), "Why Is the Pack Persuasive? The Effect of Choice Status on Perceptions of Quality", *Sociological Science*, **3**: 239–63.
Lynn, Freda B., Mark H. Walker and Colin Peterson (2016a), "Is Popular More Likeable? Choice Status by Intrinsic Appeal in an Experimental Music Market", *Social Psychology Quarterly*, **79** (2): 168–80.
Macy, Michael and Milena Tsvetkova (2015), "The Signal Importance of Noise", *Sociological Methods & Research*, **44** (2): 306–28.
Manzo, Gianluca (2013), "Educational Choices and Social Interactions: A Formal Model and a Computational Test", in G. Elisabeth Birkelund (ed.), *Class and Stratification Analysis*, vol. 30, *Comparative Social Research*, Bingley, UK: Emerald, pp. 47–100.

Freda B. Lynn and Hannah W. Espy - 9781789906851

Manzo, Gianluca (ed.) (2014), *Analytical Sociology: Actions and Networks*, Hoboken and Chichester, UK: Wiley.
Manzo, Gianluca and Delia Baldassarri (2015), "Heuristics, Interactions, and Status Hierarchies: An Agent-Based Model of Deference Exchange", *Sociological Methods & Research*, **44** (2): 329–87.
Mayr, Ernst (1961), "Cause and Effect in Biology", *Science*, **134** (3489): 1501–6.
McNamee, Stephen J. (2004), *The Meritocracy Myth*, Lanham, MD: Rowman & Littlefield.
Merton, Robert K. (1945), "Sociological Theory", *American Journal of Sociology*, **50** (6): 462–73.
Merton, Robert K. (1968), "The Matthew Effect in Science: The Reward and Communication Systems of Science Are Considered", *Science* **159** (3810): 56–63.
Merton, Robert K. and Elinor Barber (2011), *The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science*, Princeton, NJ: Princeton University Press.
Miller, Graham N.S., Freda B. Lynn and Laila I. McCloud (2021), "By Lack of Reciprocity: Positioning Historically Black Colleges and Universities in the Organizational Field of Higher Education", *The Journal of Higher Education*, **92** (2): 194–226. https://doi.org/10.1080/00221546.2020.1803026.
Monk, Ellis P. (2015), "The Cost of Color: Skin Color, Discrimination, and Health among African-Americans", *American Journal of Sociology*, **121** (2): 396–444.
Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham and Jo Handelsman (2012), "Science Faculty's Subtle Gender Biases Favor Male Students", *Proceedings of the National Academy of Sciences*, **109** (41): 16474–9.
Noble, Safiya (2018), *Algorithms of Oppression: How Search Engines Reinforce Racism*, 1st edition, New York: NYU Press.
Pager, Devah (2003), "The Mark of a Criminal Record", *American Journal of Sociology*, **108** (5): 937–75.
Pease, John, William H. Form and Joan Huber Rytina (1970), "Ideological Currents in American Stratification Literature", *The American Sociologist*, **5** (2): 127–37.
Podolny, Joel M. (1993), "A Status-Based Model of Market Competition", *American Journal of Sociology*, **98** (4): 829–72.
Podolny, Joel M. (1994), "Market Uncertainty and the Social Character of Economic Exchange", *Administrative Science Quarterly*, **39** (3): 458–83.
Podolny, Joel M. (2005), *Status Signals: A Sociological Study of Market Competition*. Princeton, NJ: Princeton University Press.
Pólos, László, Michael T. Hannan and Glenn R. Carroll (2002), "Foundations of a Theory of Social Forms", *Industrial and Corporate Change*, **11** (1): 85–115.
Reynolds, Jeremy and He Xian (2014), "Perceptions of Meritocracy in the Land of Opportunity", *Research in Social Stratification and Mobility*, **36**: 121–37.
van de Rijt, Arnout (2019), "Self-Correcting Dynamics in Social Influence Processes", *American Journal of Sociology*, **124** (5):1468–95.
Rosen, Sherwin (1981), "The Economics of Superstars", *The American Economic Review*, **71** (5): 845–58.
Ross, Catherine E. and Chia-Ling Wu (1996), "Education, Age, and the Cumulative Advantage in Health", *Journal of Health and Social Behavior*, **37** (1): 104.
Salganik, Matthew J., Peter Sheridan Dodds and Duncan J. Watts (2006), "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market", *Science*, **311** (5762): 854–6.
Salganik, Matthew J. and Duncan J. Watts (2008), "Leading the Herd Astray: An Experimental Study of Self-Fulfilling Prophecies in an Artificial Cultural Market", *Social Psychology Quarterly*, **71** (4): 338–55.
Sauder, Michael (2006), "Third Parties and Status Position: How the Characteristics of Status Systems Matter", *Theory and Society*, **35** (3): 299–321.
Sauder, Michael (2018), "Systems of Evaluation and the Matthew Effect", *Journal of Management Inquiry*, **27** (4): 362–4.
Sauder, Michael (2020), "A Sociology of Luck", *Sociological Theory*, **38** (3): 193–216.
Sauder, Michael, Freda Lynn and Joel M. Podolny (2012), "Status: Insights from Organizational Sociology", *Annual Review of Sociology*, **38** (1): 267–83.
Scarf, P. A. and M.M. Yusof (2011), "A Numerical Study of Tournament Structure and Seeding Policy for the Soccer World Cup Finals", *Statistica Neerlandica*, **65** (1): 43–57.
Sinatra, Roberta, Dashun Wang, Pierre Deville, Chaoming Song and Albert-László Barabási (2016), "Quantifying the Evolution of Individual Scientific Impact", *Science*, **354** (6312).
Wilkinson, Richard G. and Kate E. Pickett (2009), "Income Inequality and Social Dysfunction", *Annual Review of Sociology*, **35** (1): 493–511. doi: 10.1146/annurev-soc-070308-115926.
Willson, Andrea E., Kim M. Shuey and Glen H. Elder, Jr. (2007), "Cumulative Advantage Processes as Mechanisms of Inequality in Life Course Health", *American Journal of Sociology*, **112** (6): 1886–924.
Wooten, Melissa E. (2015), *In the Face of Inequality: How Black Colleges Adapt*, Albany, NY: SUNY Press.
Zuckerman, Ezra W. (1999), "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount", *American Journal of Sociology*, **104** (5): 1398–438.

Freda B. Lynn and Hannah W. Espy - 9781789906851

Zuckerman, Ezra W. (2016), "Optimal Distinctiveness Revisited: An Integrative Framework for Understanding the Balance between Differentiation and Conformity in Individual and Organizational Identities", in M.G. Pratt, M. Schultz, B.E. Ashforth and D. Ravasi (eds), *The Oxford Handbook of Organizational Identity*, Vol. 1, Oxford, UK: Oxford University Press.

Zuckerman, Harriet (1998), "Accumulation of Advantage and Disadvantage: The Theory and Its Intellectual Biography", in C. Mongardini and T. Simonetta (eds), *Robert K. Merton and Contemporary Sociology*, pp. 139–61.

Freda B. Lynn and Hannah W. Espy - 9781789906851

16. Network amplification

Filiz Garip and Mario D. Molina

INTRODUCTION

Social networks play a key role in shaping our tendency toward adopting beneficial behaviors, or avoiding risky ones, as well as our success in achieving desirable outcomes. Social scientists routinely document how having friends who excel at school (Sacerdote 2011), or those with well-paying jobs (Lin 1999; Rivera 2016), bring us closer to reaching similar results. Researchers also show how social relationships with people with healthy habits (Centola 2010) and emotional states (Fowler and Christakis 2008), or those on sensible financial tracks (Duflo and Saez 2002), make us more likely to choose similar behaviors.

Social networks can help diffuse behaviors or enable outcomes that improve our well-being. But, on the flip side, social networks can also concentrate such behaviors and outcomes in certain segments of the population and amplify differences between groups. The reason is simple. Our social relationships do not come about at random, but rather follow certain patterns. We tend to befriend people who resemble us, for instance, in socio-economic status (Marsden 1987), race (Moody 2001), and immigrant background (Smith et al. 2016), and thus, display a bias toward "homophily" (McPherson et al. 2001). We also live in societies where different characteristics, like income and race in the United States, are highly correlated. Under such "consolidation" (Blau 1977; Blau and Schwartz 1984), our friendships become segregated along many characteristics, even if we do not consider each and every characteristic in establishing our social ties (Moody 2001; Zhao and Garip 2021).

Research shows that social networks can amplify inter-group inequalities in the adoption of beneficial practices under conditions of homophily and consolidation. DiMaggio and Garip (2011) use a computational model to simulate the network-based diffusion of a practice. The higher the bias toward homophily in network formation the authors find, the greater the divergence in the adoption paths of groups defined on the basis of income, education, and race. Zhao and Garip (2021) use a slightly different modeling strategy to show that network amplification of inter-group differences depends not just on homophily, but also on consolidation. The more correlated the relevant attributes in a population are, the higher the inequality in adoption across groups.

In each of these examples, homophily or consolidation (or both) lead to social ties that are segregated along key attributes (like income or education), and this segregated structure enables the amplification of existing differences in behaviors and outcomes across groups in society. This process, however, only works if the behaviors or outcomes in question are more likely among the advantaged groups initially (e.g. the rich and educated), and if the behaviors and outcomes are subject to "network effects", that is, if they are easier to undertake or achieve when others around us are also doing them.

Recent work focuses on clarifying what "network effects" mean and suggests different mechanisms through which our alters can influence us. This development is crucial to improve our understanding of network-based inequality. In particular, we argue, research to date has

Filiz Garip and Mario D. Molina - 9781789906851

focused almost exclusively on the mechanisms for network formation (such as homophily bias) in connecting networks to inequality. Network formation is an important process to unpack in understanding network-driven differences in the distribution of behaviors and outcomes in a population. But it does not exhaust the mechanisms through which our alters shape our choices and, more generally, the extent to which networks can amplify existing differences or outcomes. Different kinds of network effects, in other words, can create different trajectories of behavior or outcome differentiation across groups that cannot be simply reduced to network formation (DiMaggio and Garip 2012).

In what follows, we first review the research on network formation, focusing particularly on work that considers homophily and consolidation as mechanisms that segregate networks along individual attributes. We then briefly discuss empirical research that suggests "network effects" in behaviors and outcomes, and provide examples of studies that connect network processes to inter-group differences in behaviors and outcomes. We then take a decidedly analytical-sociological turn and review the theoretical work laying out the mechanisms underlying network effects. Such a mechanistic approach, we argue in our concluding section, opens up a new path to understanding network amplification and its consequences for inequality.

1. NETWORK FORMATION: A PRECURSOR TO NETWORK AMPLIFICATION

Network amplification of inter-group differences depends on particular patterning of network connections. One such patterning comes from the higher likelihood of individuals to associate with those who occupy proximate social positions (Blau 1977). Our tendency to connect with people who resemble us, also known as homophily bias, implies that our experiences, activities and relationships to different social contexts (e.g. clubs, workplaces, voluntary organizations) might show strong correlations to our personal attributes (Feld 1981; McPherson and Ranger-Moore 1991; Marsden 1987; McPherson et al. 2001).

The sources of homophily are much contested in the literature. A big debate is on whether existing social groups (e.g. based on a shared attribute or a social setting) enable the formation of network ties or whether network ties eventually give rise to social groups. A useful conceptual break in thinking about social groups and social networks comes from differentiating "induced homophily" from "choice homophily" (McPherson and Smith-Lovin 1987).

Under-induced homophily, existing social groups, and their particular composition, explain similarity of connected nodes. For example, an elite school might exclusively recruit students of high-income families, and social ties within the school might inadvertently be homophilous with respect to income. Under choice homophily, friendship ties result from a deliberate preference for similarity, or avoidance of dissimilarity,[1] one that cannot be attributed to composition of the social setting alone. For example, students might select friends sharing their ethnic background even in a diverse classroom with ample opportunities to interact with other ethnic groups (Smith et al. 2016). To underline the distinction between induced and

[1] Skvoretz (2013) considers both social attraction and social repulsion as sources of homophily, suggesting that "repulsion is the driver of intergroup relations when the dimension type is ascribed like ethnicity/race and religion, while attraction is the driver of intergroup relations when the dimension type is achieved, like educational attainment" (p. 501).

Filiz Garip and Mario D. Molina - 9781789906851

choice homophily, some researchers refer to the former as "homogeneity" and reserve the term "homophily" for cases of preferential association only (DiMaggio and Garip 2011; Wimmer and Lewis 2010).

Previous research finds that induced and choice homophily do not exist in isolation. Rather, observed ties between similar people result from a combination of structural constraints and individual preferences (McPherson and Smith-Lovin 1987). Kossinets and Watts (2009) use data on electronic communications, for example, to show how induced and choice homophily work in tandem to reinforce the tendency for similar individuals to interact. As a social network evolves, similar but structurally distant individuals are brought closer together, creating a dynamic that produces strong homophily even when there is weak preference for homophilous relationships (see also Wimmer and Lewis 2010).

Homophily in network formation is important for diffusion processes because of its potential to segregate social relationships. When network ties are homophilous with respect to socio-demographic (or other) attributes, social contagion might remain stuck in certain parts of the network rather than spreading globally. Under homophily, one might observe strong empirical correlations between sociodemographic attributes and various attitudes, opinions, and behaviors, which are in fact an artifact of the network structure (DellaPosta et al. 2015). To see that, consider a setting where support for gun control is high among the educated, and where network ties are homophilous with respect to education (McPherson 2004). The observed correlation between education and gun control views in this case might be due less to a direct effect of the former on the latter and more to the ease with which the educated individuals can share common spaces and exchange ideas with similar others, thus reinforcing their own biases. Boutyline and Willer (2017), for example, show a positive correlation between political homophily and political extremism, suggesting that the clustering of the latter in certain segments of society may be less about inter-personal influence and more about preferential attachment among like-minded individuals. Consistent with this idea, Macy et al. (2019) find that political opinions depend far more on the initial conditions under which social influence takes off than on core values or ideological divisions.

Individuals who share attributes also often share multiple social spaces. Contact between different groups depends on the extent to which different social spaces overlap and, thus, bring together individuals with different attributes. Blau (1977) argued that inter-group contact, or social integration, would be low in societies with low diversity (where one group dominates others), and in societies with high consolidation of individual attributes. Consolidation makes one attribute highly predictive of the other ones, severely affecting formation of cross-cutting social circles (Blau and Schwartz 1984). For example, if income and ethnicity are highly correlated so that high-income individuals are generally white and low-income individuals are generally Latino/a, then it will be a rare event for a low-income white and a high-income Latino/a to meet, let alone share ideas and influence one another's behavior.

Low homophily and low consolidation, then, should facilitate diffusion of ideas because individuals located in different parts of the network, and with access to unique information, can still connect, and share that information, with dissimilar others. This idea, put forth by Blau (1977) and Blau and Schwartz (1984), is also reinforced by the strength-of-weak-ties hypothesis (Granovetter 1974; Burt 1992) and the small-world models (Watts and Strogatz 1998).

Yet, overlapping ties in cohesive groups (as one would observe under high homophily and high consolidation) also provide advantages in diffusion, such as more bandwidth for information sharing due to heightened social pressure and common social expectations (Aral and

Filiz Garip and Mario D. Molina - 9781789906851

Van Alstyne 2011). Using computational experiments, for example, Centola (2015) explores emergent network topologies, and diffusion outcomes, under different combinations of homophily and consolidation. Contrary to Blau and Schwartz's (1984) idea that low homophily and low consolidation lead to higher diffusion (and more social integration), Centola finds that moderate to high levels of homophily and consolidation facilitate the diffusion of norms that require reinforcement from multiple individuals and, thus, follow what Centola and Macy (2007) call a "complex contagion" process. Diffusion, in the Centola model, is facilitated by dense neighborhoods that provide exposure to multiple "infected" individuals, and by "wide bridges" that connect diverse clusters of individuals.

In support of this idea, Centola (2010) sets up an online experiment and shows that adoption of health-promoting behaviors is faster, and reaches a greater proportion of the population, when individuals are assigned a position in cluster-lattice networks (with dense and overlapping local connections) than in random networks (with little local structure). Other researchers have found either mixed support for complex contagion, for example, in diffusion of exercise behaviors (Aral and Nicolaides 2017) or showed a combination of computational models and empirical data that the theory needs to consider heterogeneous involvement of actors in the diffusion of complex technologies (Manzo et al. 2018). In particular, when early adopters that are centrally located in the network are not heavily involved in the person-to-person learning process, wide bridges can actually slow down the diffusion process by spreading uncertainty about the technology.

Research to date is, thus, unequivocal on the importance of network formation (shaped by homophily and consolidation, as well as other population parameters) for diffusion outcomes. The next step is to understand how network ties, once formed, give rise to differential diffusion trajectories across groups in a population. Before we delve into this question, we take a detour into the broader literature on network effects and briefly discuss their potential implications for inequality.

2. NETWORK EFFECTS ON BEHAVIORS AND OUTCOMES: A NECESSARY CONDITION FOR NETWORK AMPLIFICATION

There is a vast empirical literature on "network effects" where characteristics or behaviors of network alters are seen as critical to individual outcomes. Our goal here is not to offer a comprehensive account of this sprawling field, but rather to review some canonical examples and consider their implications for network-driven inequality.

Research identifies network effects in many domains, including education (Sacerdote 2011), work (Marsden and Gorman 2001), health (Pampel et al. 2010; Smith and Christakis 2008) and political participation (Campbell 2013). Empirical findings suggest positive effects of network peers on school performance (Calvó-Armengol et al. 2009), finding a job (Peterson et al. 2000), healthy behaviors (Centola 2010; Christakis and Fowler 2013; Aral and Nicolaides 2017), adoption of new technologies (Manzo et al. 2018; Keating et al. 2020), and civic participation (Lewis et al. 2013).

The extent of network effects depends on the characteristics of alters and ego. Let's focus on labor market outcomes. Individuals with ties to high-status individuals access better resources that lead to prestigious jobs (Lin et al. 1981) and promotions at work (Burt 1998). Individuals who are high-status themselves often reap higher returns from their social ties. Those with high

Filiz Garip and Mario D. Molina - 9781789906851

socio-economic status, for instance, are more likely to rely on their networks in finding a job compared to their lower-status counterparts (Smith 2007), and also to land higher-quality jobs by doing so (Lin 1999). Similarly, whites benefit more from their networks in obtaining desirable labor market outcomes compared to African Americans in the United States (Korenman and Turner 1996; Smith 2005), and men more than women (Burt 1998; Ensel 1979). (For a more detailed description of how different individuals come to reap differential returns over time, see Freda Lynn's chapter on cumulative advantage in this volume.)

These findings have clear implications for network amplification of inequality. If characteristics of alters matter in the magnitude of network effects, we can conjecture, any process that differentially distributes social ties will influence inter-group differences in outcomes. We have already reviewed one such process—homophily bias. If high-status individuals are likely to be connected among themselves, for example, the advantages of having high-status alters will be concentrated among the already advantaged (high-status) individuals, leaving those lacking status further behind. Similarly, regardless of the degree of homophily, if high-status individuals reap greater benefits from their ties, the presence of network effects in an outcome will serve to amplify existing status inequalities in that outcome.

Network amplification of group differences, although implicit in the examples above, is hard to show with data. Identification of network effects is already challenging methodologically, given the difficulties around specifying peer influences (Manski 1993, 2000; Shalizi and Thomas 2011), and separating them from the process of peer selection (Aral et al. 2009; Durlauf and Ioannides 2010; Mouw 2003). Ignoring these different pieces that contribute to the explanation of diffusion can have severe consequences for the estimation of network effects. For instance, Aral et al. (2009) study the adoption of a new technology and show that two empirical patterns often claimed to be evidence of peer influence (assortative mixing and temporal clustering) can be substantially explained by homophily. Indeed, when homophily is not accounted for, the authors find, social contagion can be overestimated by up to 700%.

If we add to these difficulties of inference, the dearth of longitudinal data on networks and behaviors required to trace group-specific trajectories, we end up with a tall order to fill, indeed. This is one reason why existing work repeatedly turns to computational models to demonstrate network-driven inequality, as we describe next.

3. NETWORKS AMPLIFICATION OF INTER-GROUP DIFFERENCES: EVIDENCE FROM COMPUTATIONAL MODELS

Empirical findings on network effects carry clear implications for social inequality. If different groups have ties to different kinds of individuals, or benefit from their ties in different ways, it is not a big leap to think that network ties will have consequences for inter-group differences. But, to understand this process systematically, it helps to study it as a unified causal chain that takes us all the way from (i) how social ties form, to (ii) how those ties shape our choices, and (iii) how those choices then aggregate into broader patterns of social inequality.

DiMaggio and Garip (2011) consider this causal chain to understand the sources of the racial divide in internet service subscription in the United States. Their argument draws on a number of stylized facts. Social networks in the United States display a high degree of homophily by education and race (Marsden 1987) and play a key role in the adoption of internet service which, like other communication technologies, becomes more valuable to us as more of our

Filiz Garip and Mario D. Molina - 9781789906851

social ties use it (Fischer 1992). The adoption of internet service is driven in part by personal resources, such as income and education and, in the United States, African Americans have lower levels of those resources, on average, compared to the whites.

These stylized facts suggest a possible explanation for the racial divide in internet adoption. First, given their higher income and education, whites are likely to be overrepresented among the early adopters of the internet in the United States. Second, given racial homophily in social ties, whites are also likely to experience stronger network effects (or what DiMaggio and Garip call "network externalities") in adoption; that is, they are likely to benefit disproportionately from their social ties. These two factors together—initial differences in adoption and differential network effects—are sufficient to give rise to an enduring racial gap in internet adoption.

To illustrate and generalize this argument, DiMaggio and Garip rely on an agent-based model, where aggregate patterns emerge from the interactions of agents following simple rules (Bruch and Atwell 2015). Instead of relying on synthetic agents, however, the authors use the respondents to the 2002 General Social Survey (GSS), and thus capture empirical distributions of (and correlations among) income, education, race, and network degree in the United States.

Their model starts by generating a network of ties among agents. Depending on the degree of homophily bias, each agent establishes a share of its target ties to socially proximate agents (in terms of race, income, and education), and the remainder of its ties randomly. Each agent has a reservation price (a price it is willing to pay for the internet service), which is an increasing function of income, education, and the share of its social contacts who have already adopted it. The price of the internet service is a declining function of the total number of adopters, reflecting economies of scale.

In this setup, an agent can adopt the internet because its reservation price has increased with more adoptions in the agent's network, or because the price of the service has dropped due to more adoptions in the population. The model runs until it reaches equilibrium by determining adopters in each time period (whose reservation price exceeds the price of the service) and updating reservation prices and the price of the service given the new adoptions.

The results from the model allow for a comparison of diffusion levels in the population and across different income, education, and race groups, under different degrees of homophily, and confirm the authors' initial intuition. As homophily increases in the network, the level of adoption in the population drops. The practice (internet service) remains concentrated among the more advantaged (the rich, educated, and white), and fails to spread to less advantaged groups (the poor, less educated, and African American). The higher the homophily, the higher the inequality in adoption.

Manzo (2013) similarly applies an agent-based model to investigate the sources of inequality in educational attainment in France. The model follows from the reasonable assumption that agents make educational choices based on their ability, expected payoffs, and choices of others in their own socio-economic group. Strikingly, the results show that the observed educational stratification in France cannot be closely reproduced in silico unless the model features group-specific network effects.

In Manzo's analysis, network effects on inequality depend both on homophily (because educational payoffs increase with the number of individuals with similar social background) and on consolidation (because ability and payoffs are positively correlated with parental education). His modeling approach also implies that increasing inter-group contact creates a zero-sum game for the individuals with different social backgrounds. That is, when groups intermingle, the low-status group benefits at the expense of the high-status group, which ends

Filiz Garip and Mario D. Molina - 9781789906851

up losing its competitive advantage, or more specifically, the "wide bridges" that enable the diffusion of better educational outcomes within this group. The results from this model, when combined, suggest that educational differentiation in France plausibly results from social influences in individual choices combined with homophily and consolidation in social ties by socio-economic status.

These studies posit similar conditions for the network amplification of inequality. To the extent that our social networks matter in adopting behaviors or achieving outcomes that help us get ahead (such as subscribing to the internet or obtaining education) and to the extent that those networks are selective with respect to attributes that also matter in our choices, social networks will amplify existing inequalities between groups. In both cases, we see that the higher the homophily in the population and the more segregated the networks (or social groups), the more divergent the trajectories of adoption across groups.

Zhao and Garip (2021) qualify these findings using a slightly different modeling strategy that varies not just homophily, but also consolidation (the correlation among different attributes) in the population. The effect of homophily on the extent of network amplification, the authors argue, is contingent on the level of consolidation. Their logic becomes clear if we consider two extreme cases. In the first case of full consolidation (where attributes are perfectly correlated), there is a single axis of differentiation. Knowing an individual's income, for instance, allows us to accurately predict their education, ethnicity, and so on. In the opposite case of no consolidation (where attributes are randomly distributed), there are multiple axes of differentiation. Knowing one attribute does not give us any information on the other attributes.

Zhao and Garip argue that the implications of homophily for network amplification will vary across these two extremes. Under full consolidation, even small degrees of homophily will be enough to concentrate adoption among those with the initial advantage (for example, the rich, educated, and white in the internet adoption case). A slight preference for similar alters in one dimension (e.g. income) will end up reinforcing homophily along other status dimensions (such as education). Under no consolidation, by contrast, high degrees of homophily will be necessary to divert advantages to groups with the initial advantage. In this case, a preference for, say, high-earning alters will not automatically yield highly educated alters, as income and education are not correlated. Put differently, the adoption advantage that comes from one's status in different dimensions (income and education, in this example) will not be consolidated.

To illustrate these ideas, Zhao and Garip use an agent-based model, introducing status differentiation into a setup used earlier by Centola (2015). Similar to DiMaggio and Garip (2011), their results suggest that status-based homophily can exacerbate inter-group inequality in the adoption of a beneficial practice (like in the internet case) when initial adopters are high-status individuals, and when adoption is subject to network effects. But different from earlier work, Zhao and Garip show that homophily drives inequality only when consolidation is moderate to high, that is, when social dimensions are correlated to a degree that the practice remains trapped in parts of the network where high-status actors cluster. Surprisingly, homophily actually alleviates inequality when consolidation is low, that is, when social dimensions are too weakly related to generate a network with overlapping ties, or "wide bridges" across which the practice can diffuse (Centola and Macy 2007).

In sum, these computational models allow researchers to trace network-driven inequality to mechanisms of network formation (such as homophily and consolidation). We now turn to an alternative direction for investigating the evolution of network-based differences: identifying the mechanisms through which networks enable the transmission of behaviors and outcomes.

Filiz Garip and Mario D. Molina - 9781789906851

4. MECHANISMS FOR NETWORK EFFECTS: THE MISSING LINK BETWEEN NETWORKS AND INEQUALITY

As mentioned, we define "network effect" simply as the effect of an alter's behavior (or outcome) on ego's choices. Network amplification of social inequality is predicated on the presence of network effects, or "social interactions" as economists typically refer to them (Manski 2000; Durlauf and Ioannides 2010). Network effects are not only difficult to identify empirically, but also hard to demarcate conceptually.

There are many mechanisms that can give rise to network effects. Indeed, researchers have come up with different typologies of mechanisms, and some have sought to characterize empirical "footprints" of each mechanism (e.g. Young 2009). We cannot do justice to this complicated literature given space limitations. But, we can point out two general questions that animate most existing categorizations of network-effect mechanisms: (1) what is it about the alters that shape ego's behavior, and (2) what is it about the ego that shifts in response?

DiMaggio and Garip (2012), for instance, focus on the first question, and define four classes of mechanisms. Each mechanism is represented via a different functional form relating alters' behavior to ego's choices. In *social contagion*, ego adopts as soon as it is exposed to a single alter depicting a behavior. Under *social learning*, ego needs to observe a certain number of adopting alters, and that threshold of adoption varies according to the riskiness of the behavior. Under *normative influence*, ego faces pressure from adopting alters in proportion to their prevalence among its immediate ties. And, finally, under *network externalities*, ego draws more benefit from a behavior the more alters adopt it. In each of these cases, ego's adoption is a function of the number or proportion of adopting alters in its network.

A similar question of alter behavior underlies Young's (2009) distinction between *contagion* and *social influence* (where ego is compelled to conform once adopting alters increase in numbers), and Rossman et al.'s (2008) definitions of *contagion*, *threshold models* (a version of social learning) and *network externalities*.

Hedström (2005) and Åberg and Hedström (2011), by contrast, focus on the second question, that is, on how the ego (or its environment) changes in response to alters' behavior. In their typology, network effects can work through ego's changing *desires* (D), *beliefs* (B), or *opportunities* (O) in relation to alters' choices (also known as the DBO theory of social interactions). Under *desire-based interactions*, ego's desires shift with alters' behavior. For example, an individual might wish to start exercising after seeing their friends criticize those who do not. Or, a person might choose to wear a mask during the COVID-19 pandemic after realizing that not doing so is stigmatized in their community. Under *belief-based interactions*, ego's beliefs about the efficacy of a behavior change after observing it among its alters. For example, an individual might take up exercising after seeing their friends improve their health by jogging regularly. Under *opportunity-based interactions*, ego's opportunities to adopt a behavior or achieve an outcome depend on the choices of its alters. For example, an individual might be more likely to find out about job opportunities in Silicon Valley if they have many friends already working there.

The DBO typology is similar to Manski's (2000) categorization of network effects. Manski identifies three channels for social interactions (economists' preferred terminology to refer to network effects): *preference interactions*, *expectation interactions*, and *constraint interactions*. These three channels correspond to the desire-based, belief-based, and opportunity-based interactions, respectively, in the Hedström classification.

Filiz Garip and Mario D. Molina - 9781789906851

The DBO and Manski classifications—although inspired by a slightly different question—also bear affinity to the categories in DiMaggio and Garip's (2012) or Young's (2009) typology. For example, *desire-based interactions* give rise to what DiMaggio and Garip call *normative influence*, while *belief-based* and *opportunity-based interactions* underlie *social learning* and *network externalities*.[2]

Regardless of the terminology and classification used, we argue, focusing on the mechanisms underlying network effects is crucial for unpacking the network amplification of inter-group differences. For example, empirically, we observe large inequalities by socio-economic status in health-improving behaviors, such as exercising, dieting, or quitting smoking, in the United States (Christakis and Fowler 2008; Pampel et al. 2010). We see relatively smaller gaps by socio-economic status in adoption of new technologies (DiMaggio and Garip 2011). The differences in relative magnitude in inequality might be connected to the differences in the mechanisms underlying network effects in the adoption of healthy habits or novel technologies.

Manzo et al. (2018) offer a perfect illustration of the importance of mechanisms in understanding variation in diffusion outcomes. Using data from Kenyan and Indian potters, along with simulations, the authors find social learning to be a key mechanism for inter-group inequality in the adoption of new technologies. Specifically, the presence of reinforcement from several early adopters activates "wide bridges" for diffusion, while absence of such reinforcement inhibits any potential influence such a network structure might have. In a similar vein, Barkoczi and Galesic (2016) argue that different social-learning strategies interact with network structure to affect group-level performance. When individuals rely on exploration and search as a strategy, for example, locally connected lattices ("inefficient networks") lead to more optimal group-level outcomes. Conversely, when individuals rely on exploitation and imitation of others as a strategy, fully connected lattices ("efficient networks") produce better group-level performance.

Identifying the mechanisms underlying inter-personal influence, as these examples illustrate, allows us not only to understand the sources of network-driven differences in society, but also to design network-based interventions to alleviate those differences (An 2015; Kim et al. 2015).

CONCLUSION

Analytical sociologists have long argued for the importance of specifying the mechanisms, or micro-level elements, underlying key relationships in the social sciences (Gambetta 1998; Hedström 2005; Hedström and Bearman 2009; Manzo 2010). This insight has made little segue into the vast "network effects" literature to date. This literature has amassed impressive empirical evidence on how social ties among individuals, and their particular patterning at the local and global levels, can be linked to various outcomes, such as educational attainment (Calvó-Armengol et al. 2009; Crosnoe et al. 2003), labor market success (Granovetter 1974; Ioannides and Loury 2004; Petersen et al. 2000), and health-improving behaviors (Christakis and Fowler 2013; Kreager and Haynie 2011). But the literature has little clarity on explaining how exactly networks matter in different instances, often reverting to a laundry list of potential mechanisms underlying alter influences on the ego.

---

[2] The so-called "social influence network theory", which highlights how ego's susceptibility to be influenced is determined by social relations, can also be related to the existing categorizations (Friedkin 1998; Friedkin and Johnsen 2011). See Childress and Friedkin (2012) for an illustrative application of the theory.

Filiz Garip and Mario D. Molina - 9781789906851

In this review, we have argued that the recent efforts to clarify the mechanisms for network effects offer a crucial path forward for understanding how networks can amplify inter-group inequality. In particular, existing work on network-driven inequality has highlighted the importance of network structure and how that structure comes to be through individual choices, like a bias toward homophily, or population-level factors, like consolidation of attributes. This literature has focused little, if at all, on how different kinds of alter effects can lead to different trajectories of behaviors or outcomes for different groups (DiMaggio and Garip 2012).

Recent attempts to develop typologies of network effects, for example, Hedström's (2005) desires-beliefs-opportunities (DBO) framework or DiMaggio and Garip's four-part categorization (social contagion, social learning, normative influence, and network externalities), bring much needed clarity to a cluttered literature. These typologies offer a natural starting point for investigating how different kinds of alter influences shape trajectories of behavioral adoption among different groups. For example, do we observe more or less inequality in healthy behaviors across socio-economic groups if those behaviors spread via peers changing our beliefs (e.g. in the benefits of exercising) or through peers changing our opportunities for adopting (e.g. by offering to exercise together)? Similarly, do we observe more or less inequality in the adoption of a technology (where peers increase its value to us by using it themselves), or in the adoption of societal norms, like vaccinating our children (where peers can actively offer rewards or sanctions)?

Future work can focus on the mechanisms underlying network effects (choosing a typology that best fits the purpose at hand) to understand how different mechanisms might drive the inequalities in the adoption of the same behavior (e.g. exercising), or to compare how different mechanisms might explain levels of inequality in the adoption of different practices (like technology and vaccination in the above example).

This direction, although fruitful, requires detailed data on individuals' behaviors and social ties over time. Furthermore, the identification of network effects, even with the requisite data, poses difficult methodological challenges. There is reason to believe that these difficulties are not insurmountable, however. Recent data sources (such as AddHealth, CILS4EU, or traces of online behaviors) offer unprecedented detailed information on the ties and choices of a large group of individuals. Similarly, recent advances in dynamic network models (Snijders and Steglich 2015) or network matching (Aral et al. 2009) offer new solutions to the identification of network effects. For these data sources and methodological developments to be of value to understanding network-driven inequality, we first need to hone in on the task of clearly specifying the mechanisms underlying network effects and theorizing about their potential impact on patterns of inequality in our particular domains of interest. This task, we argue, is one that analytical sociologists are well suited for given their orientations to sink to the micro-level to understand the macro-level phenomena, and given their particular tool kits for theory development (such as agent-based models).

REFERENCES

Åberg, Yvonne and Peter Hedström (2011), "Youth unemployment: A self-reinforcing process", pp. 201–26, in P. Demeulenaere (ed.), *Analytical Sociology and Social Mechanisms*, Cambridge, UK: Cambridge University Press.

An, Weihua (2015), "Multilevel meta network analysis with application to studying network dynamics of network interventions", *Social Networks*, **43**: 48–56.

Filiz Garip and Mario D. Molina - 9781789906851

Aral, Sinan, Lev Muchnik and Arun Sundararajan (2009), "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks", *Proceedings of the National Academy of Sciences*, **106** (51): 21544–9.
Aral, Sinan and Marshall Van Alstyne (2011), "The diversity-bandwidth trade-off", *American Journal of Sociology*, **117** (1): 90–171.
Aral, Sinan and Christos Nicolaides (2017), "Exercise contagion in a global social network", *Nature Communications*, **8**: 14753.
Barkoczi, Daniel and Mirta Galesic (2016), "Social learning strategies modify the effect of network structure on group performance", *Nature Communications*, **7**: 13109.
Blau, Peter M. (1977), *Inequality and Heterogeneity: A Primitive Theory of Social Structure*, New York: The Free Press.
Blau, Peter M. and Joseph E. Schwartz (1984), *Crosscutting Social Circles: Testing a Structural Theory of Intergroup Relations*, New York: Academic Press.
Boutyline, Andrei and Robb Willer (2017), "The social structure of political echo chambers: Variation in ideological homophily in online networks", *Political Psychology*, **38** (3): 551–69.
Bruch, Elizabeth and Jon Atwell (2015), "Agent-based models in empirical social research", *Sociological Methods & Research*, **44** (2): 186–221.
Burt, Ronald S. (1992), *Structural Holes: The Social Structure of Competition*, Cambridge, MA: Harvard University Press.
Burt, Ronald S. (1998), "The gender of social capital", *Rationality and Society*, **10** (1): 5–46.
Campbell, David E. (2013), "Social networks and political participation", *Annual Review of Political Science*, **16**: 33–48.
Calvó-Armengol, Antoni, Eleonora Patacchini and Yves Zenou (2009), "Peer effects and social networks in education", *The Review of Economic Studies*, **76** (4): 1239–67.
Centola, Damon (2010), "The spread of behavior in an online social network experiment", *Science*, **329** (5996): 1194–7.
Centola, Damon (2015), "The social origins of networks and diffusion", *American Journal of Sociology*, **120** (5): 1295–338.
Centola, Damon and Michael Macy (2007), "Complex contagions and the weakness of long ties", *American Journal of Sociology*, 113 (3): 702–34.
Childress, C. Clayton and Noah E. Friedkin (2012), "Cultural reception and production: The social construction of meaning in book clubs", *American Sociological Review*, **77** (1): 45, 68.
Christakis, Nicholas and James Fowler (2013), "Social contagion theory: Examining dynamic social networks and human behavior", *Statistics in Medicine*, **32** (4): 556–77.
Crosnoe, Robert, Shannon Cavanagh and Glen H. Elder Jr (2003), "Adolescent friendships as academic resources: The intersection of friendship, race, and school disadvantage", *Sociological Perspectives*, **46** (3): 331–52.
DellaPosta, Daniel, Yongren Shi and Michael Macy (2015), "Why do liberals drink lattes?", *American Journal of Sociology*, **120** (5): 1473–511.
DiMaggio, Paul and Filiz Garip (2011), "How network externalities can exacerbate intergroup inequality", *American Journal of Sociology*, **116** (6): 1887–933.
DiMaggio, Paul and Filiz Garip (2012), "Network effects and social inequality", *Annual Review of Sociology*, **38**: 93–118.
Duflo, Esther and Emmanuel Saez (2002), "Participation and investment decisions in a retirement plan: The influence of colleagues' choices", *Journal of Public Economics*, **85** (1): 121–48.
Durlauf, Steven N. and Yannis M. Ioannides (2010), "Social interactions", *Annual Review of Economics*, **2**: 451–78.
Ensel, Walter M. (1979), *Sex, Social Ties, and Status Attainment*. Albany, NY: SUNY-Albany Press.
Feld, Scott L. (1981), "The focused organization of social ties", *American Journal of Sociology*, **86** (5): 1015–35.
Fischer, Claude S. (1992), *America Calling: A Social History of the Telephone to 1940*, Berkeley, CA: University of California Press.
Fowler, James H. and Nicholas A. Christakis (2008), "Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study", *British Medical Journal*, **337** (a2338): 1–9.
Friedkin, Noah E. (1998), *A Structural Theory of Social Influence*, New York: Cambridge University Press.
Fiedkin, Noah E. and Eugene C. Johnsen (2011), *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*, New York: Cambridge University Press.
Gambetta, Diego (1998), "Concatenations of Mechanisms", pp. 102–24, in P. Hedström and R. Swedberg (eds), *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge, UK: Cambridge University Press.
Granovetter, Mark (1974), *Getting a Job*, Cambridge, MA: Harvard University Press.
Hedström, Peter (2005), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge, UK: Cambridge University Press.
Hedström, Peter and Peter Bearman (2009), "What is analytical sociology all about? An introductory essay", pp. 3–24, in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford, UK: Oxford University Press.

Filiz Garip and Mario D. Molina - 9781789906851

Ioannides, Yannis M. and Linda Datcher Loury (2004), "Job information networks, neighborhood effects, and inequality", *Journal of Economic Literature*, **42** (4): 1056–93.
Keating, Nancy L., A. James O'Malley, Jukka-Pekka Onnela, Stacy W. Gray and Bruce E. Landon (2020), "Association of physician peer influence with subsequent physician adoption and use of bevacizumab", *JAMA Network Open*, **3** (1): e1918586.
Kim, David A., Alison R. Hwong, Derek Stafford, D. Alex Hughes, A. James O'Malley, James H. Fowler and Nicholas A. Christakis (2015), "Social network targeting to maximise population behaviour change: A cluster randomised controlled trial", *The Lancet*, **386** (9989): 145–53.
Korenman Sanders and Susan Turner (1996), "Employment contacts and minority-white wage differences", *Industrial Relations*, **35** (1): 106–22
Kossinets, Gueorgi and Duncan J. Watts (2006), "Empirical analysis of an evolving social network", *Science*, **311** (5757): 88–90.
Kossinets, Gueorgi and Duncan J. Watts (2009), "The origins of homophily in an evolving social network", *American Journal of Sociology*, **115** (2): 405–50.
Kreager, Derek A. and Dana L. Haynie (2011), "Dangerous liaisons? Dating and drinking diffusion in adolescent peer networks", *American Sociological Review*, **76** (5): 737–63.
Lewis, Valerie A., Carol Ann MacGregor and Robert D. Putnam (2013), "Religion, networks, and neighborliness: The impact of religious social networks on civic engagement", *Social Science Research*, **42** (2): 331–46.
Lin, Nan (1999), "Social networks and status attainment", *Annual Review of Sociology*, **25** (1): 467–87.
Lin, Nan, Walter M. Ensel and John C. Vaughn (1981), "Social resources and strength of ties: Structural factors in occupational status attainment", *American Sociological Review*, **46** (4): 393–405.
Macy, Michael, Sebastian Deri, Alexander Ruch and Natali Tong (2019), "Opinion cascades and the unpredictability of partisan polarization", *Science Advances*, **5**: eaax0754.
Manski, Charles F. (1993), "Identification of endogenous social effects: The reflection problem", *The Review of Economic Studies*, **60** (3): 531–42.
Manski, Charles F. (2000), "Economic analysis of social interactions", *Journal of Economic Perspectives*, **14** (3): 115–36.
Manzo, Gianluca (2013), "Educational choices and social interactions: A formal model and a computational test", *Comparative Social Research*, **30**: 47–100.
Manzo, Gianluca (2010), "Analytical sociology and its critics", *European Journal of Sociology/Archives européennes de sociologie*, **51** (1): 129–70.
Manzo, Gianluca, Simone Gabbriellini, Valentine Roux and Freda Nkirote M'Mbogori (2018), "Complex Contagions and the Diffusion of Innovations: Evidence from a Small-N Study", *Journal of Archaeological Method and Theory*, **25** (4): 1109–54.
Marsden, Peter V. (1987), "Core Discussion Networks of Americans", *American Sociological Review*, **52** (1): 122–31.
Marsden, Peter V. and Elizabeth H. Gorman (2001), "Social networks, job changes, and recruitment", pp. 467–502 in I. Berg and A.L. Kalleberg (eds), *Sourcebook of Labor Markets: Evolving Structure and Processes*, NY: Kluwer Academic/Plenum Publishers.
McPherson, J. Miller (2004), "A Blau space primer: prolegomenon to an ecology of affiliation", *Industrial and Corporate Change*, **13** (1): 263–80.
McPherson, J. Miller and Lynn Smith-Lovin (1987), "Homophily in voluntary organizations: Status distance and the composition of face-to-face groups", *American Sociological Review*, **52** (3): 370–9.
McPherson, J. Miller, Lynn Smith-Lovin and James M. Cook (2001), "Birds of a feather: Homophily in social networks", *Annual Review of Sociology*, **27** (1): 415–44.
McPherson, J. Miller and James R. Ranger-Moore (1991), "Evolution on a dancing landscape: Organizations and networks in dynamic Blau space", *Social Forces*, **70** (1): 19–42.
Moody, James (2001), "Race, school integration, and friendship segregation in America", *American Journal of Sociology*, **107** (3): 679–716.
Mouw, Ted (2003), "Social capital and finding a job: Do contacts matter?", *American Sociological Review*, **68** (6): 868–98.
Pampel, Fred C., Patrick M. Krueger and Justin T. Denney (2010), "Socioeconomic disparities in health behaviors", *Annual Review of Sociology*, **36**: 349–70.
Petersen, Trond, Ishak Saporta and Marc-David L. Seidel (2000), "Offering a job: Meritocracy and social networks", *American Journal of Sociology*, **106** (3): 763–816.
Rivera, Lauren A. (2016), *Pedigree: How Elite Students get Elite Jobs*, Princeton, NJ: Princeton University Press.
Rossman, Gabriel, Ming Ming Chiu and Joeri M. Mol (2008), "Modeling diffusion of multiple innovations via multilevel diffusion curves: Payola in pop music radio", *Sociological Methodology*, **38** (1): 201–30.
Sacerdote, Bruce (2011), "Peer effects in education: How might they work, how big are they and how much do we know thus far?", in Eric A. Hanushek, Stephen Machin and Ludger Woessmann (eds), *Handbook of the Economics of Education*, vol. 3, pp. 249–77, San Diego, CA: Elsevier.

Filiz Garip and Mario D. Molina - 9781789906851

Shalizi, Cosma R. and Andrew C. Thomas (2011), "Homophily and contagion are generically confounded in observational social network studies", *Sociological Methods & Research*, **40** (2): 211–39.
Skvoretz, John (2013), "Diversity, integration, and social ties: Attraction versus repulsion as drivers of intra- and intergroup relations", *American Journal of Sociology*, **119** (2): 486–517.
Smith, Kirsten P. and Nicholas A. Christakis (2008), "Social networks and health", *Annual Review of Sociology*, **34**: 405–29.
Smith, Sandra Susan (2005), "'Don't put my name on it': Social capital activation and job-finding assistance among the black urban poor", *American Journal of Sociology*, **111** (1): 1–57.
Smith, Sandra Susan (2007), *Lone Pursuit: Distrust and Defensive Individualism among the Black Poor*, New York, NY: Russell Sage Foundation.
Smith, Sanne, Frank Van Tubergen, Ineke Maas and Daniel A. McFarland (2016), "Ethnic composition and friendship segregation: Differential effects for adolescent natives and immigrants", *American Journal of Sociology*, **121** (4): 1223–72.
Snijders, Tom and Christian Steglich (2015), "Representing micro-macro linkages by actor-based dynamic network models", *Sociological Methods & Research*, **44** (2): 222–71.
Watts, Duncan J. and Steven H. Strogatz (1998), "Collective dynamics of 'small-world' networks", *Nature*, **393** (6684): 440–2.
Wimmer, Andreas and Kevin Lewis (2010), "Beyond and below racial homophily: ERG models of a friendship network documented on Facebook", *American Journal of Sociology*, **116** (2): 583–642.
Young, H. Peyton (2009), "Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning", *American Economic Review*, **99** (5): 1899–924.
Zhao, Linda and Filiz Garip (2021), "Network diffusion under homophily and consolidation as a mechanism for social inequality", *Sociological Methods & Research*, **50** (3): 1150–85.

Filiz Garip and Mario D. Molina - 9781789906851

17. Complex contagions

Damon Centola

INTRODUCTION

Contemporary research on the spread of social contagions begins with a puzzle. The puzzle comes from a stark contradiction between the central claim of Malcolm Gladwell's (2000) best-selling book *The Tipping Point* and the answer to the question: How do innovative technologies—like Twitter, Facebook and Skype—and social movements—like the Arab Spring and Black Lives Matter—become successful? It is a puzzle that extends back many decades before Gladwell's book, to Mark Granovetter's (1973) famous study "*The Strength of Weak Ties*", and even further to Katz and Lazarsfeld's (1955) well-known study of "opinion leaders". Gladwell was only the most recent proponent of an influential idea that permeates the work of these important thinkers. The idea is that social innovations spread through networks by propagating just like a virus. The problem is that this widely accepted theory of social spreading does not match up with the most current data on how social innovations actually spread through social networks.

The solution to this puzzle comes from an overlooked theoretical distinction between simple contagions—like simple information and familiar ideas, which spread through single contact—and complex contagions, like changes in workplace culture or the adoption of innovative technologies, which typically require reinforcement from several peers before people are willing to adopt them (State and Adamic 2015; Steinert-Threkeld 2017; Sprague and House 2017; Mønsted et al. 2017; Traag 2016; Centola 2010, 2015). The hallmark of simple contagions is that they spread easily, like gossip or the measles. Mere exposure to an "activated" individual is typically sufficient for transmission. By contrast, complex contagions encounter resistance. Simply having contact with a single adopter is typically not enough to convince someone to adopt (Centola et al. 2007; Centola 2010, 2015).

There are four key sources of resistance that create complexity in the process of social spreading (Centola and Macy 2007; Centola 2018). Each one is a barrier to adoption. Identifying whether an innovation will encounter any or all of these barriers reveals whether an innovation will be simple or complex.

(1) Social coordination: If the value of an innovation or behavior depends on coordinating with other people who adopt it—i.e. the greater the number of adopters the more useful the innovation—then it requires social reinforcement to spread. Think of Twitter or Facebook.
(2) Legitimacy: For some behaviors, the more people who adopt them, the greater the expectation is that others will approve of the decision to adopt, and the lower the risk is of embarrassment or sanction. Think of wearing a new fashion.
(3) Credibility: It often happens that adopters act as sources of social proof for an innovation. The more people who adopt a behavior, the more likely it is that the behavior is worth the cost or the effort it takes to adopt it—and the lower the risk is of wasting time or resources on it. Think of the decision to adopt a new diet or an expensive new digital technology.

Damon Centola - 9781789906851

(4) Emotional excitement: Some innovations and behaviors are appealing only when people are emotionally energized by one another. Think of joining a protest like Black Lives Matter.

These four barriers to adoption may occur individually or together. However, all of them can be overcome by social reinforcement. For instance, innovative social media tools like Facebook and Twitter are transmitted through peer networks that provide sufficient social reinforcement to enable people to coordinate with their friends in their decisions to adopt (Ugander et al. 2012; Tool et al. 2011). The growth of political uprisings and social movements like the Arab Spring also spread via social reinforcement because people need to believe that a movement is legitimate before they will be willing to add their voice to the chorus (Steinert-Threkeld 2017). Similarly, the spread of new workplace gender norms typically requires sufficient social reinforcement to create a coordinated "critical mass" of peers that can trigger a tipping point for a new social norm (Kanter 1977; Dahlerup 1988; Centola et al. 2018). More generally, everything from the spread of memes like the Ice Bucket Challenge in the US (Sprague and House 2017) to the growth of environmental technologies like household solar panel installations in Europe (Rode and Weber 2016), are complex contagions that require social reinforcement in order to spread successfully.

The most significant consequence of the distinction between simple and complex contagions is that it transforms our understanding of how the structure of social networks impacts the dynamics of social contagion. Granovetter's (1973) superb study of the strength of weak ties shows that "weak" acquaintanceship relations tend to link randomly across a network, bridging long social distances, while "strong" intimate friendship and family relations tend to be clustered together, creating many "triadic" friend-of-friend ties. Once Granovetter developed this strikingly clear conception of the large-scale structure of social networks, he inferred that long-distance acquaintanceship ties are more valuable for the process of social diffusion than trusted friendship ties. As he put it, "whatever is to be diffused can reach a larger number of people, and traverse a greater social distance (i.e. path length), when passed through weak ties rather than strong" (1973: 1366). The crucial concept in Granovetter's analysis is *path length*: the number of steps in the shortest path from one individual to another (Solomonoff and Rapoport 1951, Rapoport and Horvath 1961). Because weak ties span long social distances, they shorten the path length between everyone in the network. The more weak ties there are, the shorter the path length between people becomes—and the faster that new ideas, innovations and social movements can spread from one person to everyone else.

But, complex contagions change everything. For a complex contagion, a single link across a social network is not sufficient to spread an innovation from one person to another. The "bridge" across the social network created by a weak tie is narrow bridge—composed of a single link. Because the spread of complex contagions requires social reinforcement from multiple contacts, an effective bridge across the network must be a wide bridge—composed of multiple ties. Wide bridges are the essential network structure needed to spread complex contagions. And, they are typically associated with strong ties rather than weak.

Strong ties naturally create wide bridges because they are clustered together, creating stable pathways for spreading complex contagions from one community to another. This is why so many innovations (from farming equipment to Twitter memberships) have historically been found to spread "spatially" or "geographically"—because spatial and geographic networks are typically composed of strong ties that form wide bridges (Centola and Macy 2007;

Damon Centola - 9781789906851

Centola 2018; Hagerstrand 1968; Toole et al. 2012; Centola 2021). But, wide bridges need not be geographically constrained. They can also span long geographic distances. On social media sites like Facebook, the emergence of "virtual" wide bridges among vast networks of peers enables social reinforcement to propel the spread of complex contagions. As you will see, social movements ranging from the Arab Spring to the growth of support for same-sex marriage are complex contagions that have spread across social media networks by using reinforcing network pathways. Remarkably, these networks bear a stronger resemblance to the strong tie network pathways through which the Civil Rights Movement took hold (McAdam and Paulsen 1993) and pre-industrial technologies spread (Hagerstrand 1968), rather than the weak tie networks through which information and viruses typically expand (Centola 2018; Centola 2021; Guilbeault et al. 2018a).

For nearly a century, network concepts such as distance, influence and centrality have been developed under the assumption that social dynamics follow the principles of simple contagion, for which each weak tie constitutes a bridge across the network. But, a narrow bridge is not a viable path for a complex contagion, which means that the distance between people (the number of steps between them in the social graph) is different for a simple contagion than for a complex contagion. So, how do we measure a network's path length? And, what does it mean for predicting the spread of social innovations? This chapter will show you how the findings on complex contagions have changed our conception of social networks, impacting everything from how we measure social distance to how we develop strategies for social change.

In the last decade, research on complex contagions has produced several key advances in the science of social networks and diffusion, including:

(1) The role of countervailing influences in diffusion
(2) The failure of "influencers" to spread innovation and the power of the periphery for initiating social change
(3) Generalized network measures of centrality, bridge width and path length
(4) Social tipping strategies for triggering tipping points in social norms
(5) The role of network bias in the failure of vital innovations and what can be done about it
(6) The value of both similarity and diversity in social networks for the spread of new ideas and beliefs
(7) Unexpected backfire effects from product awareness campaigns

This chapter will explain each of these advances. To ground the discussion, I will begin with a recent example (from State and Adamic 2015) of the distinctive dynamics of complex contagion and how they can lead to rapid, population-wide cascades of collective behavior. I will then show the implications of these dynamics for a wide range of core concepts in the study of social networks and diffusion.

1. A SURPRISING EXAMPLE

On 25 March 2013, the civil rights organization Human Rights Campaign initiated one of the largest social movements in online history. It was the same week that the US Supreme Court was hearing two cases that would decide the fate of same-sex marriage in the United States.

Damon Centola - 9781789906851

The members of the Human Rights Campaign decided to work together in a coordinated effort to encourage citizens to change their Facebook profiles to an image—a red and pink "equals sign"—that would show support for same-sex marriage.

Within a week, nearly three million people had adopted the innovation, showing unprecedented nationwide support for the initiative. If you were on Facebook at the time, you probably saw this. You may have even been part of it. The social contagion spread so quickly that magazines like the *Huffington Post* and *Adweek* wrote articles not just about HRC's campaign, but also about how quickly it was spreading. Clearly, if anything it was a case of viral diffusion, this was.

Facebook researchers Lada Adamic and Bogdan State decided to take a closer look. Poring over millions of shares, comments and likes, the scientists looked at not just the spread of the equal sign movement, but at dozens of social memes that had spread across Facebook—from "viral" photos that were widely shared and liked to "contagious" behaviors, like changing your profile to support same-sex marriage. They found something striking. While photo sharing was a simple contagion that spread very quickly through weak ties—on average spreading from person to person after only a single contact—the equality sign required more social reinforcement in order for people to adopt it. The equal sign was a complex contagion.

What was the difference between sharing a popular photo and adopting a popular profile change? While both kinds of contagions spread effectively through Facebook, and both kinds of contagions were relatively easy to adopt, they were spreading in very different ways.

State and Adamic (2015) discovered that the key to explaining the spread of the equal sign was that people needed social confirmation from multiple peers before they would believe that the movement was legitimate enough, and widely accepted enough, to support it. Unlike photo sharing, the equal sign movement spread through social reinforcement, cascading across clustered ties in the Facebook network. As State and Adamic (2015: 1741) put it:

> It is easy to see why social proof obtained from multiple sources would be necessary for many individuals to show their support for a cause they believe in. Engaging in a behavior that challenges the status quo carries inherent risks, from the minute—a quarrel with one's otherwise-thinking friends—to the life-threatening, as experienced by activists in a political movement challenging a repressive regime … [S]howing support for same-sex marriage still carries a level of perceived risk … [because] it was also likely that at least some friends and acquaintances of potential supporters of same-sex marriage held the opposing view … [Consequently,] most individuals would only change their profile picture after observing several others do it.

2. COUNTERVAILING INFLUENCES

State and Adamic's (2015) findings highlight one of the principal differences between simple and complex contagions: the influence of non-adopters. For a simple contagion, non-adopters are irrelevant. Contact with a single person who has the measles will transmit the disease regardless of how many people you know who don't have the measles. Similarly, a single person can spread information about which team won the championship game, regardless of how many people haven't found out yet.

It's different for most complex contagions.

Damon Centola - 9781789906851

Whether the decision to adopt requires coordination, credibility, legitimacy or emotional excitement, a potential adopter will not be swayed by a single adopter if everyone else they know has failed to adopt. The nonadopters are countervailing influences. They send a silent but distinct signal to potential adopters: the new behavior is not widely accepted. While some complex contagions (like establishing the credibility of a new rumor) may rely solely on having a sufficient number of contacts activated, whenever adopting a behavior requires establishing its legitimacy or coordinating with those around you, then people typically require a significant fraction of their contacts to adopt before they will. As Doug McAdam and Ronelle Paulson remarked in their study of participation in the Civil Rights Movement, "the fact that we are embedded in many relationships means that any major decision we are contemplating will likely be mediated by a significant subset of those relationships" (1993). The more non-adopter contacts that a potential adopter has, the more countervailing influences the person will confront. And, the more social reinforcement from activated peers will be needed to trigger their adoption.

3. INFLUENCERS

Countervailing influences have surprising consequences for the well-known idea of "influencers"—highly connected social stars who are able to influence lots of people simultaneously. For simple contagions, influencers can transform a small outbreak into a massive epidemic. Think of COVID-19. The disease spreads easily from person to person. If a highly connected "social star" becomes infected, that person can single-handedly spread the disease to hundreds of others (Liljeros et al. 2001; Newman et al. 2006). An influencer's remarkably large number of social contacts can make them into "super-spreaders" who can single-handedly trigger the epidemic spread of a viral contagion. This insight from epidemiology has often been generalized from the spread of viruses to the growth of social epidemics, such as the contagious spread of ideas, innovations and political movements. The journalist Malcolm Gladwell referred to the power of highly connected individuals to accelerate the spread of social contagions as "the law of the few".

However, complex contagions break this "law". When a social contagion is not merely a piece of gossip or a familiar product, but a social innovation, highly connected social stars can be roadblocks for diffusion rather than accelerants (Centola 2021b). This is because the more connected someone is, the more countervailing influences they face. For an innovation that is unpopular, contentious or socially risky—like changing your Facebook profile to support same-sex marriage—highly connected people are likely to look carefully at what others are doing before adopting the change themselves. Unlike highly connected social stars, at the "center" of the social network, people in the network periphery have only a modest number of social contacts. But, this is precisely why they are the key to social change.

In the network periphery, people do not need to contend with a sea of countervailing influences. Even a modest amount of social reinforcement from their peers can be sufficient to convince them to join a protest, since these positive signals do not need to contend with a sea of countervailing influences. In the network periphery, clusters of social change can grow—and spread—through chains of wide bridges linking moderately connected individuals. As a social movement gains momentum, it can reach a critical mass in the network periphery, at which point even highly connected social stars will be convinced to join. This is how the

Damon Centola - 9781789906851

Arab Spring protests took hold of Egypt. While there were many highly connected activists who joined the movement, the events that triggered the eruption in Egypt's Tahrir Square originated with moderately connected people who collectively mobilized to grow a revolution (Steinert-Threkeld 2017).

And it's not just social movements. The same dynamics of countervailing influences affect the spread of novel technologies. From technologies like Twitter (Toole et al. 2012) to memes like the Ice Bucket Challenge (Sprage and House 2017, these innovations are complex contagions (Centola 2021) that spread much more effectively through reinforcement in the network periphery than from highly connected influencers.

Why are influencers so reluctant early on?

Because highly connected influencers didn't become social stars by ignoring their social connections, but by paying attention to them. Influencers are keenly aware of how their behaviors are perceived, and one of the best signs of whether a contentious new idea or social movement is widely accepted is how many countervailing influences they observe in their social networks. It's the same for coordinating on a new social technology, like Twitter. For someone in the network periphery, it only takes a handful of adopters to make coordination on a new technology seem worthwhile. However, for a highly connected influencer, it takes a significant number of adopters to convince them that an unusual and unknown technology will have any coordination value for them.

For all of these reasons, social movements and novel innovations typically do not spread from highly connected influencers out to the periphery. Rather, it is the opposite. Social innovations first take hold in the network periphery and spread from there to take over the center.

4. STRUCTURE IS DYNAMICS: COMPLEX PATH LENGTH AND COMPLEX CENTRALITY

It makes intuitive sense that if a highly connected person at the center of the network adopts a new behavior, then that behavior should spread quickly and effectively to everyone else. But we just saw that this intuitive notion of spreading does not work when it comes to spreading complex contagions.

Why? Why is the periphery suddenly so important just because a social contagion is more complex? If there's more resistance to an innovation or idea, shouldn't that make influencers more important for spreading social change, rather than less? Countervailing influences are part of the explanation. But there's more to it than that.

The answer to these questions—the reason why complex contagions spread from the periphery to the center, rather than from the center to the periphery—comes from measures of network structure that reveal that the "center" of the network changes with the contagion (Guilbeault and Centola 2021).

Traditionally, the most common methods for identifying central individuals in a social network are (i) "degree centrality" (individuals with the most connections); (ii) "betweenness centrality" (individuals through which most paths must travel, going from one part of a network to another); and (iii) "eigenvector centrality" (individuals whose neighbors are highly connected). But all of these assume that a single tie between people constitutes a step in the network. It turns out that these measures of centrality—in fact all of the common measures of network topology—assume that contagions are simple. One of the most valuable findings

Damon Centola - 9781789906851

on complex contagions (from a practical point of view) has been that all these measures of network structure are defined by the complexity of the social contagion that is spreading across them.

This insight takes us back to Granovetter's observation about path length. Path length is the default measure of distance between nodes in a network used for characterizing the dynamical properties of networks. For nearly three quarters of a century, dating back at least to Somolonoff and Rapoport (1951) and Erdos and Renyi (1960), the assumption has been that a single link in a graph—for instance, a tie from A to B—is equivalent to a path between A and B. But that's only for simple contagions.

For a complex contagion, a path between A and B requires multiple ties—a wide bridge. This means that the number of "steps" in the network (or the "degrees of separation" between people) is based on the kind of contagion that is being transmitted between them. For a simple contagion, like gossip, a single tie from A to B is sufficient for transmission. For a complex contagion, like signing up for Twitter or joining the Arab Spring protests, a weak tie from A to B is not a path for transmission. For the spread of a complex contagion, the existence of a path from A to B must be composed of wide bridges at each step. Thus, even though a single weak tie may exist between A and B, the actual network path to travel from A to B may, in fact, have a path length of nine "steps", as shown in Figure 17.1.

Guilbeault and Centola (2021) refer to the classic definition of "path length" used by Granovetter and others as "simple path length", or PLs, indicating that it measures social distance in a graph based on the propagation of simple contagions. They then introduce an alternative way to measure the distance between nodes in a graph based on the propagation of complex contagions, called "complex path length", or PLc.

Complex path length measures the "steps" between nodes A and B in terms of the number of wide bridges that must be traversed to travel from A to B. This new measure of path length explains why the dynamics of social change often spread from the periphery to the center. According to this measure of social distance, the locations in the network that are most

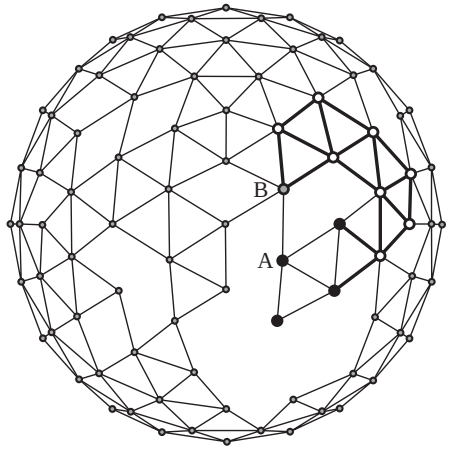


Figure 17.1 Example of a simple path from node A to node B (single tie) and a complex path from node A to node B (white nodes)

Damon Centola - 9781789906851

"central" in the network—from which complex contagions will spread most effectively—are located in the network periphery.

Based on the classic notion of simple path length, the traditional centrality measures of degree centrality, betweenness centrality and eigenvector centrality all assume that a single tie between people constitutes a step in the network. If, instead, each step in the network requires a wide bridge, the general measure (for both simple and complex contagions) of "complex centrality" identifies the neighborhoods in a social network that are most highly connected to every other neighborhood. The people in these neighborhoods are located at the intersection of more wide bridges than any other individuals in the network. Strikingly, these individuals are not very central according to standard centrality measures (Guilbeault and Centola 2021). However, when it comes to complex contagions, these network locations consistently outperform every other network location for spreading the adoption of social innovations across large and complex populations (Wang et al. 2019; Centola 2019; see also Becker 1970 and Bakshy et al. 2009).

5. SOCIAL TIPPING POINTS

Once we appreciate that measures of network centrality depend on network dynamics—that is, a person's centrality in a social diffusion process depends on whether the contagion that is spreading is simple or complex—it has enormous strategic implications. Complex centrality not only explains the apparent puzzle of why social movements spread from the periphery to the core, but also identify the particular network locations from which new social initiatives will spread most effectively. One way to apply this idea is to use complex centrality for identifying the locations in a social network that can be used to trigger tipping points.

The theory of "tipping points" in social change dates back to the 1950s when it was first used to explain emergent patterns of residential segregation. In the 1970s, Mark Granovetter (1978) and Thomas Schelling (1978) generalized these ideas to study critical mass dynamics in collective behavior—from the rapid growth of social movements to the surprising dominance of new intellectual traditions. Simultaneously, Rosabeth Moss Kanter (1977) discovered that the logic of tipping points also applied to the dynamics of changing gender norms in organizations. Today, tipping dynamics have been found among across a wide variety of social change events (Centola et al. 2018). For instance, the Black Lives Matter movement was supported by well below 50 per cent of Americans in 2014, but it jumped to above 75 per cent of Americans by 2020 (Cohn and Quealy 2020).

This rapid change in public opinion can be understood by the dynamics of complex contagions. The reason the movement was initially resisted is also the reason it grew so rapidly. In 2014, people needed to believe in the legitimacy of the movement before they would join in. The movement initially took hold in the periphery, expanding through networks of wide bridges. But, once it grew large enough to establish its broad legitimacy, a tipping point was triggered that quickly saw the mushrooming of public support.

So, how should an activist target a social network to trigger a tipping point?

The most popular strategy has been to target the influencers. Following the metaphor of a virus, the assumption has been that the most highly connected people—that is, the most "central" actors based on traditional measures of centrality—are the most effective individuals for spreading social innovations. But, countervailing influences, complex path length and

Damon Centola - 9781789906851

complex centrality show otherwise. The network locations with the greatest complex centrality are within the network periphery by classical standards, but they are the most central locations for activing a critical mass that can trigger a tipping point for social change.

6. NETWORK BIAS

An unexpected implication of these findings on centrality is what they reveal about how bias takes hold in social networks, and what it means for spreading social innovations. It has become customary to think of bias as a feature of individuals: some people are highly biased and others less so. A common assumption is that information is the cure. All that is required is to give people access to correct information and their biases will be eliminated.

But this dream of reason did not take social networks into account.

Decades of failed policy efforts—from public health initiatives like promoting MMR vaccinations and COVID-19 precautions, to scientific initiatives to spread warnings about climate change—have found that simply disseminating accurate scientific information is insufficient to change people's beliefs and behaviors. The problem is not that people don't have access to information. Rather, it's that new information is filtered through people's social networks—which shape how they interpret it and what they ultimately believe.

The mistake is to think of social networks merely as pipes—conduits for the passage of diseases and information. That may be true for simple contagions, but not for complex ones. For complex contagions, social networks are also prisms. Depending on how we receive new information, who we hear it from, and how the community around us perceives it, we may either become readily convinced by a new idea or staunchly opposed to it. The networks around us can trigger motivated reasoning (Festinger 1957) that reduces important new scientific information about public health or climate change to nothing more than a "hoax", while misinformation that confirms a community's biases spreads effectively and is widely believed.

The reason for this frustrating asymmetry between accurate information and misinformation is that ideas that reinforce a community's existing beliefs are simple contagions. They are easy to understand and easy to spread. Within communities that share a set of beliefs, highly connected influencers at the center of the conversation can easily spread misinformation that plays to a group's biases (Becker et al. 2017; Becker et al. 2019; Guilbeault et al. 2018b; Guilbeault and Centola 2020b). By contrast, contentious or novel ideas that challenge a group's biases are complex contagions. These kinds of opinions face strong opposition, and thus are not likely to emerge from highly connected individuals facing a sea of countervailing influences (Samuelson and Zeckhauser 1988; Steinert-Threkeld 2017; Centola 2018, 2021).

When COVID-19 hit the US in the spring of 2020, public health officials used the media to disseminate information about new health behaviors that could slow the spread of the deadly new virus. Chief among them were wearing face masks and practicing social distancing. At first, it appeared as if these new behaviors would spread like the virus—reaching everyone and becoming a worldwide phenomenon just like the novel coronavirus. But, as the months progressed, and the weather warmed, some communities embraced the new practices while others actively opposed them. Invisible lines of social distinction (such as party affiliation) were visibly reflected in people's behavior. Some groups followed the social norms of wearing face masks and social distancing—and even tried to enforce them—while other groups aggressively flouted these norms. Face masks became ground zero for a new kind of epidemiological culture war.

Damon Centola - 9781789906851

Why would something as basic as protecting yourself (and others) from the spread of a deadly virus become so socially complex and politically divisive? The answer comes from the simplicity and complexity of these social contagions. In communities in which there was widespread distrust of government propaganda, public health interventions were viewed as an infringement on personal rights. Public health advice was not seen as credible, and was resisted. People instead coordinated on a set of behaviors that reinforced their beliefs about the virus (and the precautions it required) being a "hoax". Misinformation that confirmed these beliefs was a simple contagion and spread effectively in these communities. By contrast, COVID-19 precautions (and, on the horizon, COVID-19 vaccinations) were complex contagions, which encountered staunch opposition.

However, in communities that viewed government advice as a public service designed to protect people, there was widespread conformity with face masks and social distancing. Public health advisories were carefully followed, and anger grew over other people's willing disregard for safety protocols while grocery shopping, walking in public parks and standing in line for ice cream.

Public spaces became battlegrounds for conflicting social norms. A public health behavior that was a simple contagion for one community was a complex contagion for others. The problem that was preventing the new behavior from spreading was that community belonging and social (and political) identity played an important part in determining who was seen a relevant source of social influence, and who was not.

7. THE POWER OF SIMILARITY AND DIVERSITY

When people come into contact with a new idea or behavior they don't simply observe the number of people adopting. They also see who those people are. Are they like me or different? Are they like each other or diverse? When people were deciding whether to wear face masks, they were keenly aware of whether the people who were wearing them were part of their social group, or not. The more conspicuous the differences, the more clearly the behavior became viewed as an "out-group" behavior—not just different, but oppositional.

The spread of new ideas, innovations and beliefs from one group to another depends on wide bridges between the groups. But it also depends on people seeing the adopters as relevant. For many sociologists, it has been common to think of similarity as the defining feature of relevance (Granovetter and Soong 1988; McPherson et al. 2001; Lazarsfeld and Merton 1954; Centola 2011). That is true in some cases, but there are also situations in which diversity can be more effective than similarity for creating relevance.

The explanation for whether similarity or diversity creates relevance brings us back to the four key barriers to adoption: social coordination, legitimacy, credibility and emotional excitement. Determining which of these factors are the primary barriers to adoption reveals which features of the adopters—similarity or diversity—will be effective for creating relevant sources of social influence.

When credibility is required—for instance regarding the accuracy of new public health information about COVID-19, the safety of vaccinations or new scientific reports about the dangers of global warming—people look for social proof from similar peers. Social reinforcement from adopters is most effective when it indicates that the new behavior or idea is accepted by your "in-group" (Davis and Greve 1997; Centola and van de Rijt 2015; Howe and Monin 2017).

Damon Centola - 9781789906851

Similarly, when emotional excitement is required—for instance in recruitment to join the Arab Spring revolution, or in joining in to run the streets in celebration of victory by your home town sports team—once again, similarity with the adopters increases their ability to provide effective social reinforcement. Seeing others "like you" join in the excitement is how social effervescence grows (Centola 2021).

However, when it comes to establishing the broad legitimacy of an activity or social movement—for instance changing your Facebook profile to show support for same-sex marriage—then diverse sources of social reinforcement will be more effective. In State and Adamic's (2015) study of the spread of the equal sign, they found that people's sense of social risk could only be overcome when they saw multiple adopters from diverse social groups.

It's easy to see why. On Facebook in 2013, people were connected to their friends from high school and college, their colleagues at work, their parents, their grandparents and their neighbors. Showing support for a contentious political movement could trigger aggressive commentary on your Facebook page, and potentially even alienate or offend members of your community. However, once people saw that support for the movement was being taken up by people from diverse social quarters—their grandparents, their high school friends and their colleagues—they were convinced of the broad legitimacy of the movement, which reduced the social risk of joining and triggered widespread adoption.

8. BACKFIRE EFFECTS

For people who are interested not only in understanding the dynamics of social contagions, but also in spreading social innovations of their own, they need a way to navigate all of these subtleties of social diffusion—simple and complex contagions; simple and complex path length; various notions of centrality; asymmetry in the influence of the influencers; network bias; and varying roles similarity and diversity in creating relevance. One of the most common solutions to this problem is to take a "kitchen sink" approach.

Instead of trying to get too specific, why not just do everything? Hire influencers and target the periphery. Buy television advertisements and try to create a critical mass. Build narrow bridges and wide ones. Foster influence from similar peers and diverse ones. Surely, more is better. Even if an innovation is complex, using the strategies of simple contagion can't hurt, right?

Wrong.

I call this the *kitchen sink fallacy*. The fallacy is that any solution, even if it's not quite the right one for your particular problem, is better than no solution. Unfortunately, that's not what the science of complex contagions shows us. To the contrary, aggressively applying strategies that are effective for spreading simple contagions can backfire, actively preventing the uptake of complex contagions (Centola 2021).

Take Google+.

In 2011, Google launched Google+, its fourth and final foray in the social media market. Their strategy was to spread their social technology through a massive awareness campaign. Google's search engine had already dominated the market for nearly a decade, and Google's webmail service, Gmail, had recently overtaken AOL as the dominant webmail service worldwide. They had also acquired the world's most popular video-sharing site, YouTube.

Damon Centola - 9781789906851

They wanted to use each of these services to spread worldwide adoption of Google+. It would be the epitome of "cross-promotion". Done right, they could create widespread acceptance of their new social technology and mount a serious challenge to the established industry leader, Facebook. That's exactly what they set out to do. By 2013, every person who used a Google service was automatically (in many cases involuntarily) associated with a Google+ account. Almost instantly, everyone had Google+, and everyone knew that everyone had it. Google had achieved worldwide spread of awareness: everyone knew what Google+ was.

But, the problem was that no one was using it. People were still using Facebook. And, this caused a backfire effect. Google's massive informational awareness campaign—forcibly associating many millions of people with Google+ accounts—inadvertently created worldwide social proof against their own technology (Centola 2021). By 2016, it was clear that Google+ had failed. By 2019, Google+ closed its doors permanently. The explanation for Google's backfire effect takes us all the way back to where we started: the problem of countervailing influences.

It's one thing for a technology not to be adopted because people don't know about it yet. In that case, non-adopters are not a very strong signal about the product itself. They simply indicate that it's not popular and therefore not yet easy to coordinate on. These are the kind of countervailing influences that can prevent adoption early on. However, this can change, and the product can easily become more attractive if more people start using it.

The situation with Google+ was different. Because of Google's enormously successful awareness campaign, everyone not only knew about Google+, they also knew that everyone else knew about it. The lack of use of Google+ was incredibly conspicuous. It meant that every person who continued to use Facebook was offering an explicit rejection of Google+; these were significant countervailing influences. People were left with the need to explain why no one was using the new technology and, very soon, people concluded that the product's lack of use must be due to some flaw in the technology.

What Google overlooked in its remarkably successful awareness campaign is that a social networking platform—unlike a search engine portal or a webmail client—requires social coordination. People need to adopt together. The challenge of moving people from a powerful incumbent like Facebook to a new platform cannot be bested simply by spreading informational awareness. Success requires spreading the complex contagion of coordinated social action. In this sense, spreading a new social media technology is more like the spread of the Arab Spring than it is like the spread of gossip.

CONCLUSION

Since the 1950s (Lazarsfeld and Merton 1954; Katz and Lazarsfeld 1955) sociologists have appreciated that social networks are the crucial link between individual behavior and social outcomes, with Granovetter's (1973, 1978) groundbreaking work in the 1970s establishing the prototype for the later development of network science as a field of sustained inquiry in analytical sociology. Building on this tradition, the study of complex contagions has deepened and broadened the exploration of the micro-macro dynamics of social behavior—from new findings on innovation diffusion in the archeological record (Manzo et al. 2018) to new insights into how consolidation and homophily affect social integration and inequality (Blau and Schwartz 1984; Centola 2015; Zhao and Garip *forthcoming*). The study of complex

Damon Centola - 9781789906851

contagions has uncovered new terrain in social networks research that includes not just the structure of interpersonal interactions and social influence, but also the ways in which social networks mediate the expansion of different kinds of social innovations, as well as how differing kinds of social contagions interact with each other, often producing highly counterintuitive collective outcomes (Zhang and Centola 2019; Campbell and Salathe 2013; Su et al. 2016).

The success of a new scientific area can be measured by its intellectual productivity. How many existing puzzles does it solve and how many new ones does it uncover? How many new insights and conceptual advances does it produce, and how much new scientific exploration does it foster? By these lights, the study of complex contagions is a highly productive field of inquiry, with researchers currently exploring new applications to substantive topics, such as political influence, bullying, gender relations and immigration, as well as new theoretical studies exploring the role of influencers, the relevance of diversity, the meaning of centrality, the emergence of bias, the dynamics of inequality and the possibility of backfire effects. New research on how simple and complex contagions spread through society continues to reveal the hidden pathways through which social networks and personal influences give rise to unanticipated collective behavior.

REFERENCES

Bakshy, Etan, Brian Kerrer and Lada Adamic (2009), "Social Influence and the Diffusion of User-Created Content", *Proceedings of the 10th ACM Conference on Electronic Commerce*: 325–34.

Becker, Joshua, Devon Brackbill, and Damon Centola (2017), "Network dynamics of social influence in the wisdom of crowds", *Proceedings of the National Academy of Sciences*, **114** (26): E5070–E5076.

Becker, Joshua, Ethan Porter, and Damon Centola (2019), "The wisdom of partisan crowds", *Proceedings of the National Academy of Sciences*, **116** (22): 10717–22.

Becker, Marshall H. (April 1970), "Sociometric location and innovativeness: Reformulation and extension of the diffusion model", *American Sociological Review*, **35** (2): 267–82.

Blau, Peter M. and Joseph E. Schwartz (1984), *Crosscutting Social Circles*, Orlando, FL: Academic Press.

Campbell, E. and M. Salathé (2013), "Complex social contagion makes networks more vulnerable to disease outbreaks", *Sci. Rep*., **3**: 1905.

Centola, Damon (2010), "The spread of behavior in an online social network experiment", *Science*, **329** (5996): 1194–7.

Centola, Damon (2011), "An experimental study of homophily in the adoption of health behavior", *Science*, **334** (6060): 1269–72.

Centola, Damon (2015), "The social origins of networks and diffusion", *American Journal of Sociology*, **120** (5): 1295–338.

Centola, Damon (2018), *How Behavior Spreads*, Princeton, NJ: Princeton University Press.

Centola, Damon (2019), "Influential networks", *Nature Human Behavior*, **3** (1–2).

Centola, Damon (2021), *Change: How to Make Big Things Happen*, New York: Little, Brown & Co.

Centola, Damon and Arnout van de Rijt (2015), "Choosing your network: Social preferences in an online health community", *Social Science & Medicine*, **125**: 19–31.

Centola, Damon, Joshua Becker, Devon Brackbill and Andrea Baronchelli (2018), "Experimental evidence for tipping points in social convention", *Science*, **360** (6393): 1116–19.

Centola, Damon and Michael Macy (2007), "Complex contagions and the weakness of long ties", *American Journal of Sociology*, **113** (3): 702–34.

Centola, Damon, Victor Eguiluz and Michael Macy (2007), "Cascade dynamics of multiplex propagation", *Physica A*, **374**: 449–56.

Cohn, Nate and Kevin Quealy, (2020), "How public opinion has moved on Black Lives Matter", *The New York Times*, June 10.

Dahlerup, Drude (1988), "From a small to a large minority: Women in Scandinavian politics," *Scandinavian Political Studies*, **11** (4): 275–97.

Davis, Gerald and Henrich R. Greve (1997), "Corporate elite networks and governance changes in the 1980s", *The American Journal of Sociology*, **103** (1): 1–37.

Damon Centola - 9781789906851

Erdös, P. and A. Rényi, A. (1960), "On the evolution of random graphs", *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, **5**: 17–61.
Festinger, Leon (1957), *A Theory of Cognitive Dissonance*, Stanford, CA: Stanford University Press.
Gladwell, Malcolm (2000), *The Tipping Point: How Little Things Can Make a Big Difference*, Boston, MA: Little, Brown.
Granovetter, Mark (1973), "The strength of weak ties", *American Journal of Sociology*, **78** (6): 1360–80.
Granovetter, Mark (1978), "Threshold models of collective behavior", *American Journal of Sociology* **83** (6): 1420–43.
Granovetter, M. and Roland Soong (1988), "Threshold models of diversity: Chinese restaurants, residential segregation, and the spiral of silence", *Sociological Methodology*, **18**: 69–104. https://doi.org/10.2307/271045.
Guilbeault, Douglas and Damon Centola (2020), "Networked collective intelligence improves dissemination of scientific information regarding smoking risks", *PLoS ONE*, **15** (2): e0227813.
Guilbeault, Douglas and Damon Centola (2021), "Topological measures for identifying and predicting the spread of complex contagions", *Nat Commun*, **12**: 4430.
Guilbeault, Douglas, Joshua Becker and Damon Centola (2018b), "Social learning and partisan bias in the interpretation of climate trends", *Proceedings of the National Academy of Sciences*, **115** (39): 9714–19.
Guilbeault, Douglas, Joshua Becker and Damon Centola (2018a), "Complex contagions: A decade in review", in Yong Yeol Ahn and Sune Lehmann (eds), *Complex Spreading Phenomena in Social Syst*ems, New York: Springer Nature.
Hagerstrand, Torsten (1968), *Innovation Diffusion as a Spatial Process*, Chicago: University of Chicago Press.
Howe, Lauren and Benoît Monin (2017), "Healthier than thou? 'Practicing what you preach' backfires by increasing anticipated devaluation", *Journal of Personality and Social Psychology*, **112** (5): 735.
Kanter, Rosabeth Moss (1977), *Men and Women of the Corporation*, New York: Basic Books.
Katz, Elihu (1957), "The two-step flow of communication: An up-to-date report on an hypothesis", *Public Opinion Quarterly*, **21**: 61–78.
Katz, Elihu and Paul Lazarsfeld (1955), *Personal Influence*, New York: The Free Press.
Lazarsfeld, Paul F. and Robert K. Merton (1954), "Friendship as a social process: A substantive and methodological analysis", *Freedom and Control in Modern Society*, **18** (1): 18–66.
Liljeros, Fredrik, Cristofer Edling, Luis Amaral, H. Eugene Stanley and Yvonne Aberg (2001), "The web of human sexual contacts", *Nature*, **411**, (6840): 907–8.
Manzo, G., S. Gabbriellini, V. Roux and F. Nkirote M'Mbogori (2018), "Complex contagions and the diffusion of innovations: Evidence from a small-N study", *Journal of Archaeological Method and Theory*, **25** (4): 1109–54.
McAdam, Douglas and Ronnelle Paulsen (1993), "Specifying the relationship between social ties and activism," *American Journal of Sociology* **99 (3)**: 640–67.
McPherson, Miller, Lynn Smith-Lovin and James M. Cook. (2001), "Birds of a feather: Homophily in social networks", *Annual Review of Sociology*, **27**: 415–44.
Mønsted, Bjarke, Piotr Sapieżyński, Emilio Ferrara and Sune Lehmann (2017), "Evidence of complex contagion of information in social media: An experiment using Twitter bots", *PLoS ONE*, **12** (9): e0184148.
Newman, Mark (2010), *Networks: An Introduction*, Oxford: Oxford University Press.
Newman, Mark, Albert-László Barabási and Duncan J. Watts (2006), *The Structure and Dynamics of Networks*, Princeton, NJ: Princeton University Press.
Padgett, John F. and Christopher K. Ansell (1993), "Robust action and the rise of the Medici, 1400–1434", *American Journal of Sociol*ogy, **98** (6): 1259–319.
Page, Scott E. (2007), *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton, NJ: Princeton University Press.
Rapoport, Anatol and William Horvath (1961), "A study of a large sociogram", *Behavioral Science*, **6** (4): 279.
Reagans, Ray and Bill McEvily (2003), "Network structure and knowledge transfer: The effects of cohesion and range", *Administrative Science Quarterly*, **48** (2): 240–67.
Rode, Johannes and Alexander Weber (2016), "Does localized imitation drive technology adoption? A Case Study on Rooftop Photovoltaic Systems in Germany", *Journal of Environmental Economics and Management*, **78**: 38–48.
Romero, Daniel, Brendan Meeder, and Jon Kleinberg (2011), "Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter", *Proceedings of the 20th International Conference on World Wide Web*: 695–704.
Samuelson, William and Richard Zeckhauser (1988), "Status quo bias in decision making", *Journal of Risk and Uncertainty*, **1**: 7–59.
Schelling, Thomas (1978), *Micromotives and Macrobehavior*, New York: Norton.
Solomonoff, Ray and Anatol Rapoport (1951), "Connectivity of random nets", *Mathematical Biophysics*, **13**: 107–17.
Sprague, Daniel and Thomas House (2017), "Evidence for complex contagion models of social contagion from observational data", *PLoS ONE*, **12** (7): e0180802.
State, Bogdan and Lada Adamic (2015), "The diffusion of support in an online social movement: Evidence from the Adoption of Equal-Sign Profile Pictures", *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*: 1741–50.

Damon Centola - 9781789906851

Steinert-Threkeld, Zachary C. (2017), "Spontaneous collective action: Peripheral mobilization during the Arab Spring", *American Political Science Review*, **111** (2): 379–403.
Su, Yuan, X. Zhang, L. Liu, S. Song and B. Fang (2016), "Understanding information interactions in diffusion: An evolutionary game-theoretic perspective", *Front. Comput. Sci*., **10** (3): 518–31.
Toole, Jameson L., Meeyoung Cha, and Marta C. González (2012), "Modeling the adoption of innovations in the presence of geographic and media Influences", *PLoS One*, **7** (1): e29528.
Traag, Vincent (2016), "Complex contagion of campaign donations", *PLoS One*, **11** (4): e0153539.
Ugander, Johan, Lars Backstromb, Cameron Marlowb and Jon Kleinberg (2012), "Structural diversity in social contagion", *Proceedings of the National Academy of Sciences*, **109** (16): 5962–6.
Wang, Xioachen, Yueheng Lan and Jinghua Xiao (2019), "Anomalous structure and dynamics in news diffusion among heterogeneous individuals", *Nature Human Behavior*, **3**: 1–10.
Zhao, L. and Filiz Garip (forthcoming), "Network diffusion under homophily and consolidation as a mechanism for social inequality", *Sociological Methods and Research*.
Zhang, Jingwen and Damon Centola (2019), "Social Networks and Health: New Developments in Diffusion, Online and Offline", *Annual Review of Sociology*, **45** (1): 91–109.

Damon Centola - 9781789906851

18. Archival data

Katharina Burgdorf and Henning Hillmann[1]

1. INTRODUCTION

These days, we are witnessing an ever-increasing scope and availability of archival data sources, just waiting to be included in the social scientist's arsenal of research instruments. Not too long ago, it used to be a daunting task for the uninitiated sociologist to identify suitable depositories, then to plough through seemingly endless piles of loose-leaf manuscripts in the quest for some nugget of useful information. Next, the scattered pieces of information had to be collected, cleaned and coded into a database, and this often had to be done by hand. In other words, archives used to be viewed as the quintessential source for "hard-to-get" data, at least for the single researcher who lacked the proverbial army of research assistants. The situation has changed radically with the rise of readily accessible digital archives that cover salient information from important institutions over long stretches of time. For example, the entire career trajectories of every single member of the German Parliament since 1949 are now available to help us understand both the strategic career moves of political elites and the organizational mechanisms at work in parliamentary systems (Bailer et al. 2018). Already available for some time, the RUSCORP database contains all for-profit corporations founded in the Russian Empire from the reign of Peter the Great to the Great War, enabling social scientists to study serial entrepreneurship under the fragile institutional conditions of transition economies (Owen 1992; Hillmann and Aven 2011). Long-run developments in the self-selection of migrants to the United States and their prospects in the host country can be traced systematically by linking historical Ellis Island arrival records and U.S. Census data, both accessible in digital form (Abramitzky and Boustan 2017). As we show in more detail later, the digital archive of the Internet Movie Database (IMDb) offers multiple opportunities to study professional career systems, the emergence of new movements and creative styles, and long-run changes in the billion-dollar film industry. These are just a few examples that we are familiar with from our own work, and many more such repositories could be cited.

What remains unchanged, though, is that these new possibilities usually emerge from unstructured qualitative data, in the sense of non-numerical information that was recorded and kept first and foremost for administrative reasons, and thus is rarely ready-made for the purposes of quantitative social science research.[2] It likewise remains prudent for social scientists to understand the selectivity built into these sources: why did administrators consider some information worth recording, and with what purpose in mind, whereas other data were left off of the record? Once they were established, why did some depositories withstand the vagaries

[1] We are grateful to Christina Gathmann and Gianluca Manzo for their suggestions and comments on an earlier draft, and to Lea Cohausz for her excellent research assistance.

[2] Needless to say, unstructured data deposited in archives are not limited to administrative sources, and may include all variants of found, oral and trace data.

Katharina Burgdorf and Henning Hillmann - 9781789906851

of history whereas others did not stand a chance of survival? Was it the result of randomness or choice? (Esch 1985).

In this chapter, we show that, despite these challenges, seemingly "hard-to-get" archival data are often ideally suited to answer the kinds of research questions that analytical sociologists ask. We further distinguish traditional tangible archives—repositories of manuscripts that need to be accessed in person—and new digital archives such as the IMDb that are often associated with the idea of "big data". Drawing both on exemplary studies in analytical sociology and our own experience with archival data sources, we consider where the latter offer new solutions for old problems and avenues for addressing important new questions.

2. THE ANALYTICAL STYLE OF ANALYTICAL SOCIOLOGY

Much like other social scientists, analytical sociologists strive for precise explanations of social phenomena. Their work usually begins with an empirical puzzle: the observation of behaviors and patterns of social relationships that are interesting and unexpected because they are counterintuitive and challenge received wisdom from extant studies. Solving such empirical puzzles motivates their research. What we can discern in the work of analytical sociologists is a distinctive preference for causal explanations that rest on becoming narratives: where do markets come from (White 1981)? What are the origins of new organizational forms and innovations (Padgett and Powell 2012)? How do interest groups coalesce into new political movements and parties (Hillmann 2008)? The emphasis here is on dynamics and the emergence of collective outcomes across social space and over time. It contrasts with existence proofs that merely describe the presence of markets, organizations, or movements, and not how they come into being. To explain the emergence, reproduction, and subsequent change of collective outcomes, analytical sociologists seek to identify the micromechanisms whereby people's decisions and interactions accumulate into larger and often complex macrostructures. Ideally, regularities in the patterns of micro-macro relationships will be identified such that explanatory theories arising from one set of cases will help us understand other cases in similar empirical settings.[3]

Though not necessarily labeled as analytical sociology, research in this tradition has contributed important insights about the social world. Among others, Schelling (1978) simulated the process of how individual preferences and choices drive ethnic segregation in urban neighborhoods. Similarly, Granovetter's (1978) threshold model demonstrates how individual decisions to participate in collective action are contingent on the revealed preferences of others to participate. For example, individual people will join a protest movement only when they believe that the benefits of participation outweigh its risks, and this usually happens once a sufficient number of others have joined. Naturally, risk aversion varies from person to person,

[3] Of course, we do not mean to suggest that this analytical strategy is exclusive to analytical sociology, but that it is typical of its proponents. Likewise, we should stress that micromechanisms are not always considered necessary to explain macrostructures. Peter Blau's (1977) later work on inequality and heterogeneity as well as the population ecology tradition in organizational theory (Hannan and Carroll 2000) are just two prominent examples of approaches that seek to explain variation in macrostructure (e.g. the emergence of new organizational forms) with reference to macrolevel mechanisms (e.g. density dependence, or niche crowding in markets).

Katharina Burgdorf and Henning Hillmann - 9781789906851

but the likelihood of collective action to arise can be inferred from the distribution of personal thresholds in a population.

While some of the early analytical studies have relied on simulated data, an increasing body of literature uses diverse empirical data to analyze the interplay between micro-level behaviors and macro-structure properties. Padgett and Ansell (1993) exploit a rich portfolio of archival administrative data, such as information on taxes, business ties, marriage, and friendship records to study the rise of the Medici in Renaissance Florence. They map the evolving social network of the Florentine elite to show how the Medici's advantageous structural position, spanning multiple social categories, facilitated their capacity to mobilize political supporters. In a similar vein, Padgett and McLean (2006) trace how the organizational logic of forming partnerships in business rippled through Renaissance Florentine society and emerged as a model for social organization in multiple spheres of life, from arranging marriages between families to building alliances across political factions.

In another case of long-run macro-level change, Stovel et al. (1996) draw on personnel records of successive job positions held by the employees of Lloyds Bank between 1890 and 1970. These unique data allow the authors to trace and compare the career sequences of all employees at this important financial institution. In particular, they demonstrate how increasing mobility in individual careers combined with organizational growth to yield the distinctly modern idea of an achievement-based profession.

Hedström et al. (2000) take data from the central archives of the Swedish labor movement to examine how the movement emerged and diffused across different regions. Analyzing the travel routes of individual agitators between 1894–1911, they show how a dense information network between distant regions emerged as an unintended by-product of political agitators' visits of previously unconnected places.

Stark and Vedres (2006) collected data from official courts of registry and analyzed ownership histories of more than 1500 Hungarian enterprises between 1987 and 2001. The time frame captures the transition period from planned economy to market coordination in Hungary. For most companies, this seismic transformation also implied a shift from state to private ownership, sometimes coupled with foreign investment. Similar to the Lloyds Bank study mentioned above, the authors identify the sequential pathways taken by Hungarian firms as they navigate the challenges and opportunities of different ownership patterns. Once again, a macrostructure—here the post-socialist landscape of industrial organization—emerges from microlevel behavior—here the formation of recombinant property networks within individual firms.

Erikson and Bearman (2006) use ships' logs, factory correspondence, and journals to compile a dataset of 4572 voyages undertaken by English East India Company (EIC) traders between 1601 and 1833. They show that the commercial expansion of the EIC was driven primarily by the activities of private traders who pursued their interests under the auspices of the company. When EIC captains and traders in the East Indies delayed their return voyages to England and knowingly diverted their ships from prescribed trade routes and sailed into new ports, they seemingly engaged in opportunistic behavior. Yet, in doing so they explored new trading opportunities, linked previously unexploited ports into a dense commercial network, and exchanged valuable information about these new opportunities. Hence, the collective outcome of these seemingly opportunistic private activities formed the foundation for the success of the EIC.

Other recent studies have used digital data, or historical sources that have already been translated into digital repositories to address sociological questions. Among others, Lutter (2015) employs data from the IMDb to examine the mechanisms that drive

Katharina Burgdorf and Henning Hillmann - 9781789906851

gender inequality in the American film industry. Analyzing collaboration networks of 97 657 film actors in 36 9099 films produced between 1929 and 2010, he finds that women's embeddedness in highly cohesive networks negatively affects their career success. Elsewhere, Lena and Pachucki (2013) use data on musical samples and Billboard rankings between 1979 and 1999, collected from both published and online sources, to study the process of artistic legitimation in the case of rap music. Analyzing the relationship between artistic status and popularity they reveal the micro-mechanisms that produced status arrays in an emerging artistic field.

3. TYPICAL DATA STRUCTURES

As diverse as the topics they explore might be, these exemplary studies all draw their inferences from a particular type of data structure. If one of the hallmarks of studies in analytical sociology is their focus on emergence and social dynamics, then longitudinal data are required to capture the development of behaviors and collective outcomes in their temporal order. Ideally, such data would extend over long stretches of time—years rather than days, and decades rather than years—for the sociologist to trace long-run behavioral and institutional change. As mentioned earlier, the other hallmark of analytical sociology appears to be an emphasis on identifying micromechanisms that are deemed responsible for macrostructural changes. More often than not, the term "micro" points to individual behaviors, and the term "mechanism" implies moving parts that set something larger in motion once they work together (think of the combination of pedals, crankset, chain, sprockets, and wheelset that translate the muscle power of a rider into the forward movement of a bicycle). Once more, the Erikson and Bearman (2006) article on the EIC serves as a case in point: here, the emergent macrostructure was the EIC's global network of overseas commerce that embedded all ports along the major trade routes. The micromechanism that brought the network about worked through the exploits of individual traders who pursued their own private interests despite being agents of the company, and who exchanged valuable information about their newfound trading opportunities.

The typical data structure that satisfies the requirements of both hallmarks is inherently sequential: it records repeated observations of individual actors, their choices and behaviors over time, and in some cases, even across geographic or social space. In the case of early modern trade networks, the repeated observations are of captains and their vessels navigating the seas, from one port to the next, over successive sailing seasons. In another setting, they are yearly observations of Hungarian firms that cycle through shifting patterns in their ownership networks as they try to adapt to the challenges of a post-socialist economy. In yet another context, we may observe, year after year, how each employee within a banking firm moves from one job to the next, up or down the status ladder, over the course of his/her career (including, of course, observations of those who never get promoted and stay in their current position). Social scientists well versed in the literature on social stratification will readily recognize such sequential data structures from studies of occupational mobility, especially within internal labor markets. The link to research on careers and occupational mobility is probably no accident, as the concept of a career as a series of promotions of individual workers (read: micro-level choices) has long been considered central for understanding how firms organize hierarchies of job titles (read: macrostructure), and set wages accordingly (White 1970; Baker et al. 1994a, 1994b). Indeed, for lack of a better term, we may characterize the entire class of data structures favored and required by many analytical sociologists as career-type data.

Katharina Burgdorf and Henning Hillmann - 9781789906851

While career-type datasets are ideally suited for the aims of analytical sociologists, they usually have to be built from extant records that were collected and kept for purposes other than social science research. Archives figure prominently among such extant sources. In the remainder of this chapter, we address some of the advantages and obstacles one faces when archival sources are employed as evidence for the causal link between micromotives and macrostructures.

4. USING ARCHIVAL SOURCES

Archives come in various flavors that appeal to different tastes, depending on one's research interests. Here we suggest that it is useful to distinguish between, on the one hand, traditional tangible archives that primarily house collections of manuscripts in their original paper-based form and, on the other hand, digital archives that store all kinds of found and administrative information in digital format (Salganik 2019). Both variants of archival repositories have benefits and costs of their own, as we discuss in the two illustrative studies that follow. We share insights from two of our own research projects that draw on evidence from tangible and digital archives, respectively.

4.1 Using Tangible Archives: French Privateering Networks

Our first case to illustrate the use of traditional archival sources is Hillmann's (2021) work on French privateering networks in the long eighteenth century, from the 1680s through the onset of the French Revolution. The book shows how partnerships in privateering ventures fashioned a network organization—early modern in spirit, yet modern in form—that held together the merchant community of Saint-Malo, a key port in the French Atlantic economy. Because the empirical setting is removed in place and time from most contemporary contexts, a quick summary of the study's narrative and findings should be helpful.

Overseas expansion during the early modern period opened up new trade routes for European merchants to explore. At the same time, intense rivalries pitted the rising European states against each other in recurrent wars. Focusing on the French experience, and the port of Saint-Malo in particular, the book tells the story how local merchant communities were able to secure their hold onto established trades, to seize new opportunities, and to withstand the threats of armed conflict. Much of the answer lies in the economic activity of privateering, known as the *guerre de course* in French. In wartimes, rulers commissioned shipowning traders to fit out vessels as privateers, or *corsaires* in French, whose mission it was to raid enemy shipping. Within a mercantilist worldview, this state-licensed private war-at-sea aligned the interests of local elites and the royal government. Locally, within Saint-Malo, the partnerships that merchant elites formed in their privateering ventures gave rise to a cohesive network that held their community together amidst outside conflicts (note again the link between micro behavior and the resulting macrostructure). Privateering also served as an ideal entry point for young members of the merchantry to launch their careers in maritime commerce. The book thus reveals the local organizational foundations of early modern capitalist development and the beginnings of prescribed career trajectories. However, as the French Revolution approached, growing rifts divided the Saint-Malo elite, illustrating the limits of the integrative power of privateering enterprise. Combining a close reading of historical sources with quantitative network analysis, the book places conflict at the center stage of network formation.

Katharina Burgdorf and Henning Hillmann - 9781789906851

Central to the findings that Hillmann presents in the book is the complete set of partnership declarations (*actes de société*) for ventures of maritime trade and privateering that were registered with the local office of the Admiralty at Saint-Malo. Today, the original manuscripts can be found in the series 9B165–176 and 9B587 of the Admiralty records, supplemented by partnership contracts in the law office records of notaries Pitot (4E5277–5278), Le Roy (4E5343) and Morel (4E5364), all of which are preserved at the Archives départementales d'Ille-et-Vilaine in the city of Rennes, France. For the more than one hundred years, between 1681 and 1792, covered in the study, the declarations for a total of 3250 voyage partnerships, joining 6298 individual shipowners, captains and investors, have survived, and all of them are included in the empirical analyses of the book.

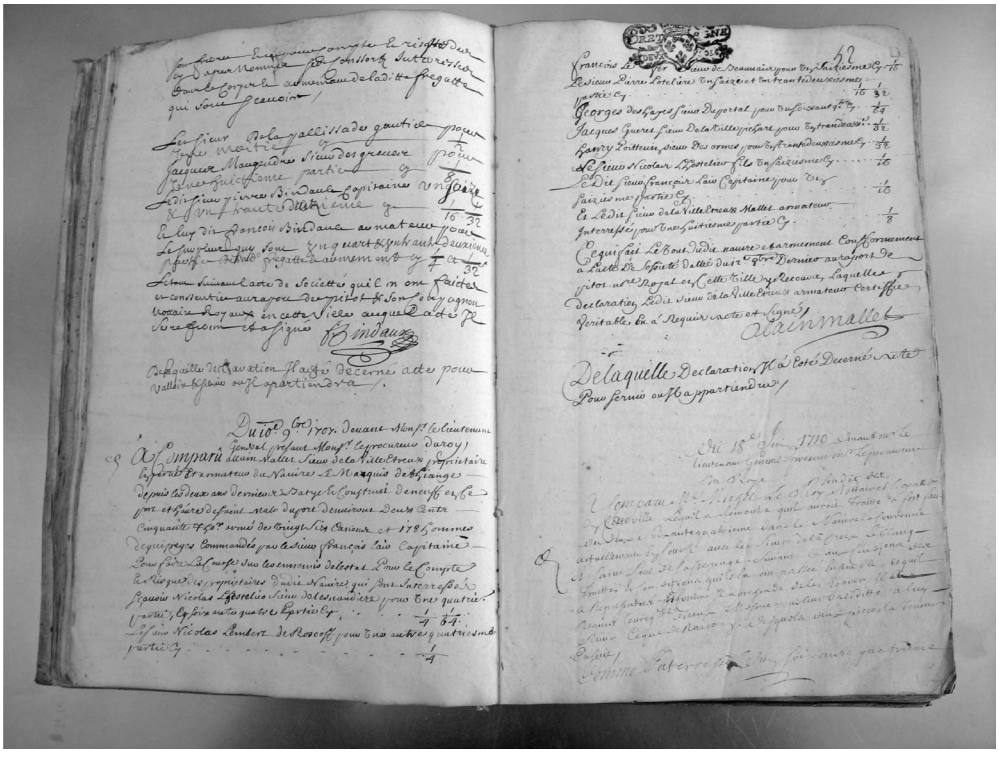
The partnership records provide a unique and extraordinary wealth of information for an in-depth understanding of the economic history of Saint-Malo as well as the local organizational foundations of the French Atlantic economy at large. What makes the source so valuable is first its sheer scope, both in terms of the long-time series and the range of different ventures it covers, from privateering enterprise, North Atlantic cod fishing and coastal trading to commerce with the Spanish and Portuguese American colonies, slave trading and long-distance trade with India and China. As such, it comes close to a census of the registered partnerships in maritime commerce organized at Saint-Malo in 1681–1792. With respect to the individual traders included, the records provide precisely the kind of career-type data we described earlier: they document the serial investments of merchant traders in various venture partnerships over time, much like the careers of employees moving from one job to the next.

Figure 18.1 shows an example of the manuscript records retrieved from the archives, in this case the contract for a privateering partnership, registered on 10 November 1707. It is the contractual agreement for the 250-ton corsair *Le Marquis de Thianges*, built in Saint-Malo, armed with 26 cannons and manned with 178 seamen (the entry starts at the bottom left and continues on the top right). A certain Alain Mallet, Sieur de la Villestreux was one of the ship's owners and *armateur* ("proprietaire En partie Et armateur du Navire"). Under the command of Captain François Lair, the *Marquis de Thianges* was fitted out for a privateering cruise against the enemies of the state ("Commandé par le sieur François Lair Capitaine Pour faire La Course sur les ennemis delestat"). What follows is the list of partners ("Interressés"), beginning with Nicolas L'Hostelier, Sieur delesnaudiere who held 1/4 and 1/64 of the shares in the venture, and ending with the captain François Lair, who contributed capital for 1/16, and the *armateur* Mallet, Sieur de la Villestreux himself for 1/8 of the shares. The closing statement declares that this summary document conforms to the *acte de société*, as notarized by the city's royal notary Pitot on 12 October, and that said Sieur de la Villestreux is indeed the rightful *armateur.* As such, he also signed the document.

Drawing on these qualitative archival data, Hillmann first codes affiliation (two-mode) networks that connect individual merchants to the ventures of trading voyages in which they participated. Since he is interested in the development of social cohesion versus fragmentation within the merchant community over time, Hillmann splits these networks into six discrete periods.[4] The period-specific networks are sensitive to historically meaningful events and at the same time based on the observed duration between the subsequent ventures that each merchant

[4] Similar to other longitudinal network data that stretch over a long time span, finding the right balance between historical detail and a sufficiently broad view of social structures tends to be a thorny issue. For details on the rationale behind the chosen periodization, see Hillmann (2021, chapter 6).

Katharina Burgdorf and Henning Hillmann - 9781789906851



Source: Archives départementales d'Ille-et-Vilaine, cote 9B168. Printed with permission.

Figure 18.1 Manuscript page of a venture partnership agreement

participated in. For each period, Hillmann constructs a binary merchant-by-voyage venture matrix with merchants arrayed in rows and voyage ventures in columns. In the first period (1681–1697), for example, the matrix cross-tabulates 723 individual merchants with 239 trading voyages. Each cell in this matrix reports if a merchant participated in the *société* of business partners who fitted the vessel for its trading or privateering voyage. Hillmann then transforms each of the six period-specific matrices into a symmetric merchant-by-merchant matrix and a corresponding symmetric voyage venture-by-voyage venture matrix. Within each resulting merchant-by-merchant network, pairs of merchants are linked to the extent that they were business partners in the same trade or privateering ventures. Entries are equal to zero in the absence of such partnership relations.[5] Their mirror image can be found in the voyage-by-voyage networks, which document the extent of interlocks between trading ventures, that is, the number of merchants that each pair of ventures has in common as business partners. With

[5] In practice, it is computationally more efficient to use edge lists instead of matrices to transform such large networks. In this case, an edge list consists of successive pairs of merchants and the voyages they invested in. Using the information contained in the archival records, various attributes can be attached to each edge linking a nodal pair, such as the type of trade, the value of the share that the merchant held in the voyage, and the date when the investment was made (which usually, though not always, equals the onset of a partnership).

Katharina Burgdorf and Henning Hillmann - 9781789906851

this relational database in hand, one can then turn to the quantitative network analysis of what used to be qualitative archival information—as Hillmann does in the book to show how individual preferences in the selection of business partners and in the formation of their ventures gave rise to a cohesive merchant community (see in particular Hillmann 2021, chapters 6–7).

Of course, there are also caveats when it comes to the use of such archival data. Without any claim to comprehensiveness, let us briefly consider two data-related challenges that routinely confront the user of archival sources. One recurrent issue is the correct identification of individual people. In this and other historical settings, families had the unfortunate (at least for the social scientist) habit of naming their offspring after their ancestors. What appear to be multiple observations of a single person—say, one Nicolas Magon—may in fact refer to grandfather and grandson, or father and son, or uncle and nephew, or two cousins, all of whom may have been alive and well at the time the document in question was set up. The problem only increases with members of large families that extend into several branches. Because repeated observations of the same persons are so central to the kinds of data structures that analytical sociologists rely on, it is important to disentangle their unique identities. Likewise, inferences drawn from a network dataset are meaningful only if any two nodes within these networks actually do refer to distinct actors. Social science history tends to deal with dead people, however, and they make poor informants when it comes to verifying their own identity. One often helpful option is to triangulate one's primary data with additional sources. In this case, we are fortunate enough that comprehensive parish registers of baptisms, marriages, and burials have been compiled for Saint-Malo and its surrounding towns (Paris-Jallobert 2004). Likewise, Hillmann was able to collect the complete poll tax rolls for selected years from the departmental and municipal archives. The tax records list each inhabitant, sorted by household and place of residence (the quarter), and the amount of taxes he or she was supposed to pay.[6] Hence, a careful inspection and comparison of names across the three sources is rather effective in identifying the identity of each person included in the dataset.

Another frequent concern is the potential for elite selectivity in administrative records of the past, favoring the illustrious at the expense of the middling sorts and lower strata. A central argument Hillmann makes in the book is that privateering partnerships in particular were pivotal for community cohesion, precisely because they attracted the interest and investments of all strata within the local merchant society. The argument is certainly difficult to sustain if the underlying data select, primarily the wealthiest strata, such that evidence on the participation of those with more modest means is absent. Even though the book is a study of a merchant elite in a local place, there is no a priori reason to assume that elites necessarily formed a single bloc without any stratification in terms of social status, wealth and political influence. Hence, the important question here is to what extent the partnership records reflect the stratification within the merchant community. Fortunately, the poll tax records mentioned earlier can be used again to establish the socio-economic standing of the merchant investors who came together in their privateering and trading partnerships. As expected, the matching of entries in the partnership and tax records reveals that variation in their financial abilities translated to variation in the number and types of trading enterprises that merchants were able to invest in. Low-cost ventures primarily attracted those of lesser means, whereas ambitious long-distance endeavors remained the preserve of Saint-Malo's most potent entrepreneurs. Crucially, few

[6] Archives départementales d'Ille-et-Vilaine, C4070, C4086, C4098, C4259, C4262; Archives municipales de Saint-Malo, CC41.

Katharina Burgdorf and Henning Hillmann - 9781789906851

if any barriers to entry existed in privateering partnerships such that purchasing shares was apparently possible for members of all merchant strata, even for those traders whose range of commitments was restricted by their limited funds (for details, see Hillmann 2021, chapter 3).

More could be said about the quantitative analysis of qualitative historical data, concerning both its potential and limits (see the list of references). The main purpose here is to offer non-specialist readers who share the research agenda of analytical sociology a first glimpse of what can be done with archival sources, and what empirical strategies may be adopted to solve problems that one is likely to encounter with such data.

4.2 Using Digital Archives: The Rise of New Hollywood Filmmakers

Recall that by digital archives we mean internet sources that provide access to found or administrative information that lends itself to a career-type data structure. This class of sources may also include digitized versions of tangible archives, such as the Harvard Digital Collection, the German Emigrants Database, or the UNESCO Digital Library. Likewise included are less institutionalized online sources, such as Wikipedia, web of science, archINFORM, or the IMDb. The IMDb, for example, provides detailed information on the careers of all actresses, directors, cinematographers and producers active in the global film industry. They may not be explicitly designed for sociological research, yet these data offer rich empirical material to study important social processes such as status emergence, reputation diffusion and job mobility.

Although the opportunities opened by the recent increase in digital depositories can stimulate innovative research, they also imply methodological challenges that affect established notions of data generation, sampling and measurement. Here we document some of these methodological challenges that Burgdorf encountered in a work-in-progress on the rise of New Hollywood filmmakers, using data from the IMDb (Burgdorf 2020). In this project, she investigates the social dynamics that underlie collective identity formation of the New Hollywood movement in the 1960s and 1970s. She asks how New Hollywood filmmakers constructed their novel auteur identity, which implied an artistic understanding of filmmaking and challenged the previous studio identity of the Golden Age of Hollywood. Burgdorf tests to what extent New Hollywood filmmakers relied on social ties and symbolic ties for the construction of their identity. To address this question empirically, she analyzed the collaboration and co-citation networks of more than 15 000 American filmmakers that participated in 52 000 films between 1930 and 2000. Modeling these networks over time, Burgdorf examined both the temporal changes in the macro structure and the micro processes that drove the emergence of the macro structure.

Using data from digital archives, such as IMDb data, for scientific research raises two essential questions. First, why do we observe the data that we observe? This question concerns the principles that underlie the generation of the observable information and points to potential selection biases caused, for example, by popularity and age. Second, how could these data be collected? This question implies more pragmatic aspects and considers hands-on tools and techniques for extracting or scraping relevant information from the website.

With respect to the data-generating process, it is important to be aware of and find solutions for potential selection biases. IMDb aims to include all film projects and persons involved over the whole film historical time. The database is created through open-source contributions and checked and adjusted by IMDb-employed editors. The validity of IMDb data has been

Katharina Burgdorf and Henning Hillmann - 9781789906851

addressed and approved by several sociological studies (e.g. Wei 2019; Lutter 2015; Cattani et al. 2014; Rossman et al. 2010; Zuckerman et al. 2003). Though striving for completeness, it may miss information.

Missing data are unproblematic as long as they do not follow a systematic pattern, implying that the information is missing completely at random. For archival data, however, it has been found that missingness is often correlated with popularity or age (Martin 2017). Take the example of Martin Scorsese. Due to his popularity, information on him may be more comprehensive compared to his less well-known colleagues. The listed information on Martin Scorsese captures even some of his early experimental works (e.g. Vesuvius VI released in 1959) that were post hoc defined as career relevant. This careful collection of early-career achievements may not apply to less popular filmmakers. Similarly, large-scale film productions that turn out as popular successes are likely to contain comprehensive information on cast, crew and revenues, while independent low-budget productions may be captured in less detail.

Another potential source of incomplete information is age. It is known from historical research that selectivity increases with the age of documents. This may yield an unknown number of early Hollywood films that are not listed on IMDb. Being aware of these potential biases is crucial to ensure "that our theory of selectivity doesn't assume the very thing that we are trying to test" (Martin 2017: 207). For example, we may hypothesize that there were overall fewer collaborations during the early days of Hollywood in the 1910s as compared to later periods. If we indeed find evidence for this claim, this may actually result from missing crew information of early films rather than a true lack of early collaborations.

Selectivity checks are one way to rule out potential biases in the data and, thus, later results. These are not straightforward and depend on the data at hand. To tackle selectivity through popularity, for example, one could ask if a film's complete documentation (meaning that all variables are listed and include perfect information) is correlated with its popularity. For the assessment of completeness, one could use specific variables such as box office information, coded as a dummy (0 = no box office information available, 1 = box office information available). For the assessment of popularity, one could use the number of reviews and the ranking position. As an illustration, Scorsese's *Vesuvius VI* received 145 votes in total and includes no box office information. *Taxi Driver* received 680815 votes and includes detailed information on box office revenues. In this illustrative case, we observe a positive relationship between completeness of information and popularity. Obviously, it is not possible to check the absence of films in the same way. To find out about overall missing films, one would have to consider additional sources, such as filmmakers' biographies and studio or university archives.

Concerning the pragmatic aspects of the data-generating process, retrieving data from digital archives is mostly cheap. In contrast to tangible archives, digital archives are easier to access and, given the right tools, information can be directly extracted. Hence, it is useful to be trained in computational tools and techniques. If one does not want to end up in a tedious loop of manual copying and pasting, the systematic collection of digital archival data requires scraping skills. When attempting to scrape a website, such as the IMDb, one needs to look at the HTML code running in the background that is structuring the content so that the algorithm one programs can use it to capture and store relevant information. For most purposes, only some parts of the code are relevant, for example, information on collaborations between filmmakers, thus, filmmakers and films. The identification and download of this specific part in the code requires the XPath. It can be found using Firefox's pre-installed Inspector

Katharina Burgdorf and Henning Hillmann - 9781789906851

or Chrome's SelectorGadget extension. Once one has identified the correct location, one can explore the meta-information stored in the code. In the case of IMDb, every movie and every involved person holds a unique identifier that is hidden in the HTML code. The URLs leading to films on IMDb include the films' unique identifiers, so by iterating through the numbers one can automatically scrape many different films. R's "rvest" (Wickham 2016) and Python's "beautifulsoup" (Richardson 2007) are packages that are easy to use to download and extract the required information. Before setting up the scraping procedure, it is important to check if the website allows scraping and if it has a limit which requires downloading at a moderate rate.[7]

The availability of novel computational methods, cheap storage facilities and access to technologies allows turning unstructured information from various archival sources into analyzable data sets. In order to save time and computational energy, it is crucial to think about the sampling procedure before starting to collect the data. Otherwise one ends up collecting, storing and analyzing huge amounts of data that are not needed to answer sociological questions. One should pay particular attention to the boundaries of the sample concerning place, time and relevant agents.

As Burgdorf (2020) seeks to understand the American auteur movement, she only includes American directors and writers. She uses two sampling approaches to capture the New Hollywood movement. In the first approach, she focuses only on collaborations and co-citations between New Hollywood filmmakers as taken from an IMDb list of 61 filmmakers. As the New Hollywood movement took place throughout the 1960s and 70s, Burgdorf only considers those films that were released between 1950 and 2000 to capture the time before, during and after the movement. This approach facilitates a precise view on the core of the movement, covering its emergence, prime and decline. It implies, however, a major drawback with regard to sample selectivity. The list includes only those filmmakers that were ex post declared as particularly salient and are now popularly known as New Hollywood filmmakers. It ignores less visible filmmakers that, nevertheless, may have contributed to the movement's identity formation.

Therefore, in the second sampling approach, Burgdorf includes all American filmmakers who participated in at least two movies during 1930 and 2000. About 70 per cent of filmmakers drop out immediately after the first film and thus are unlikely to contribute to the movement's identity formation. The remaining sample still yields more than 15 000 filmmakers. Analyzing their collaboration and co-citation networks allows capturing the social dynamics in the whole field over an extensive time period.

The measurement procedure with found data is hardly comparable to, for example, pre-designed item batteries in surveys as they are not initially collected with regard to testing theoretical concepts. Using these data requires thorough reflection on existing analytical tools. While collective identity is often conceptualized and measured through expressed identification with a group's norms and values, Burgdorf conceptualizes collective identity in relational terms, i.e. through collaboration and co-citation networks, and analyzes the temporal evolvement of both networks from a macro- and micro-sociological perspective. In a first

[7] Helpful sources for scraping include the following links:
For an introduction to HTML, XML, and rvest using Chrome:
https://towardsdatascience.com/tidy-web-scraping-in-r-tutorial-and-resources-ac9f72b4fe47
For an introduction with Python's beautifulsoup:
https://towardsdatascience.com/how-to-web-scrape-with-python-in-4-minutes-bc49186a8460

Katharina Burgdorf and Henning Hillmann - 9781789906851

step, to capture the changes in the macro structure, she measures network modularity as it is particularly suitable to operationalize cohesion in large-scale networks (Moody and Coleman 2015). Specifically, she measures the temporal changes in modularity using three-year moving windows between 1930 and 2000. In a second step, Burgdorf focuses on the underlying micro mechanisms and studies the tie generation processes that underlie the emergence of the collaboration, respectively co-citation network. She applies exponential random graph models to measure to what extent collective identity is created by generalizable social mechanisms that have been found in other contexts, such as preferential attachment, transitive closure, and homophily (Lusher et al. 2013).

Unexplored administrative data from digital sources offer invaluable opportunities for novel insights. Given that many data sources do not come with standardized procedures on how to develop sample schemes, cope with selectivity issues, and develop and apply appropriate measures, it is critical to carefully reflect on these aspects. For a first orientation, one should check how other studies that used similar data structures have operationalized the concepts one uses. Be creative, reflexive and transparent on the single steps you take, and avoid getting lost in the depth of data exploration. As John Levi Martin (2017) summarizes his thoughts on dealing with data from documents: "Accept that gift horse. But inspect its mouth carefully" (p. 209).

5. CONCLUDING REMARKS

Missing data may be the one problem that haunts social scientists working with archival sources the most. The problem has two sides. One side has to do with the idea that more information yields greater precision in one's inferences. The other has to do with the possibility that previously unknown data may threaten to undermine what is currently known, by adding perspectives and altering the meaning of cases (Bearman et al. 1999). Consider the question who the Phoenicians of the ancient Mediterranean world really were, in our modern sense of a united, self-identifying people (Quinn 2017). It is apparently a question that fired not only the popular imagination, but also motivated a significant amount of research by scholars of antiquity. Finding a definitive answer has proved to be so difficult because *references to* the existence of the Phoenicians may be plentiful, but they come almost exclusively from Graeco-Roman sources. In contrast, any direct evidence *from* self-identifying Phoenicians themselves is nearly non-existent. However, as a recent reviewer of Quinn's book put it so eloquently, "an absence of evidence is not necessarily evidence of absence" (Cioffi 2019: 15).

What if such evidence would suddenly become available right at our fingertips, and not only sporadically, but in spades? For the social science historians who keep lamenting the lacunae in their source materials this sure must be a dream come true. In the beginning of this chapter we have alluded to the recent availability of very large datasets that were previously hard to access for interested scholars. We are of course not the first to point to the new ubiquity of data. Among others, Bearman (2015) has argued that the big data revolution opens plenty of promising opportunities for a new approach in social science history, one that seeks to unearth the mechanisms whereby multiple long-run processes—the advance of market economies, nation-states, class-based stratification, and the very idea of occupational mobility—interlock with each other to bring about large-scale societal change such as the rise of modernity. Whereas studying just a single of these processes used to be a daunting task,

Katharina Burgdorf and Henning Hillmann - 9781789906851

so-called big data seem to have revolutionized our ability to do these complex transformations justice in their entirety.

Drawing on Arthur Danto's philosophy of history, Bearman has less hope that the ubiquity of data will benefit traditional social science research that is wedded to writing narrative sentences. Danto (1985) wondered if the missing data problem in historical research could be solved if we imagine an ideal chronicle that records all events the way they happen, at the very moment they happen. With an ideal chronicle, worries about selected samples of the past would dissipate because we would have access to the population of events just as they occur. However, Danto (1985: 158) concluded that an ideal chronicle is of little use to historians because their work consists primarily of writing *meaningful* narrative sentences such as the following: "the author of *Principia* was born at Woolethorpe on Christmas Day, 1642". The author in question is Isaac Newton, and the *Principia Mathematica,* first published in 1687, is now recognized as a foundational work of modern science. Today we find this sentence meaningful because it relates the birth of Newton to his later fame as a pioneer of modern physics. It is a narrative. On that Christmas Day in 1642, the ideal chronicle could not have recorded this sentence because it could not have known at the time of birth that Newton would write this book in the future. The reference to forthcoming events was unknown, and a woman giving birth to a child in a rural English hamlet was all there was to record.

It is our sense that the causal explanations favored by most practicing analytical sociologists fall squarely into the class of narrative sentences.[8] Individual choices form the beginning of the explanatory narrative, followed by a mid-section where these choices lead to selected interactions among people, or groups of people. Toward the ending, these interactions accumulate into, sometimes unintended, collective outcomes that eventually feed back to inform individual choices as preferences in the next round. To the extent that big data are not unlike Danto's ideal chronicle, it is hard to see how access to massive new data sources will reshape the explanatory enterprise of analytical sociology beyond an improvement in measurement precision.

Another challenge may be even more salient. Perhaps because explanatory arguments put forward on the basis of archival data seem to be particularly vulnerable to missing information, they are most at risk to be lured by the sheer mass of newly available data and subscribe to excessive pattern recognition where algorithms relegate theory to the back bench. As Eric Leifer (1992) alerted us nearly thirty years ago, there will always be some discrepancy between theory and data, but with such a substantial increase in the number of data points as we are witnessing now, just about any theory is bound to be falsified. If anything, then, the recent ubiquity of information should lead us to entertain a return to strong theory that necessarily has to deny data in order to create previously inconceivable objects of research, objects that allow us to understand the past, and hence the present in a new and different light (Leifer 1992). As a case in point, consider again the quest for Phoenician identity we mentioned above. Here, recent work, motivated by strong theory, suggests that the idea of "Phoenicia" was primarily an invention of the Greeks and the Romans to mark their competitors in the Mediterranean world (Quinn 2017). Opening a window to such a radical new view is called scientific progress.

---

[8] It is probably no accident that *Analytical Narratives* is the title of a highly influential volume written by prominent political economists who seem sympathetic to the program of an analytical sociology (Bates et al. 1999).

Katharina Burgdorf and Henning Hillmann - 9781789906851

REFERENCES

Abramitzky, R. and L. Boustan (2017), "Immigration in American history", *Journal of Economic Literature*, **55** (4): 1311–45.

Bailer, S., E. Frech, N. Goet, S. Hug, O. Huwyler, P. Manowand Tomas Turner-Zwinkels (2018), *Parliamentary Careers in Comparison Project Data Set*, Codebook v. 3.3.02.

Baker, G., M. Gibbs and B. Holmstrom (1994a), "The internal economics of the firm: Evidence from personnel data", *The Quarterly Journal of Economics*, **109** (4): 881–919.

Baker, G., M. Gibbs and B. Holmstrom. (1994b), "The wage policy of a firm", *The Quarterly Journal of Economics*, **109** (4): 921–55.

Bates, R.H., A. Greif, M. Levi, J.-L. Rosenthal and B.R. Weingast (1999), *Analytic Narratives* Princeton, NJ: Princeton University Press.

Bearman, P. (2015), "Big Data and historical social science", *Big Data & Society* (July–December): 1–5.

Bearman, P., R. Faris and J. Moody (1999) "Blocking the future: New solutions for old Problems in historical social science", *Social Science History*, **23** (4): 501–33.

Blau, P.M. (1977), *Inequality and Heterogeneity: A Primitive Theory of Social Structure*, New York: The Free Press.

Burgdorf, K. (2020), "The emergence of auteur filmmaking: Identity formation through symbolic networks", Unpublished Ph.D dissertation proposal, University of Mannheim.

Cattani, G., S. Ferriani and P.D. Allison (2014), "Insiders, outsiders, and the struggle for consecration in cultural fields: A core-periphery perspective", *American Sociological Review*, **79** (2), 258–81.

Cioffi, R.L. (2019), "A palm tree, a colour and a mythical bird", *London Review of Books*, **41** (January).

Danto, A.C. (1985), *Narration and Knowledge*, New York: Columbia University Press.

Erikson, E. and P. Bearman (2006), "Malfeasance and the foundations for global trade: The structure of English trade in the East Indies, 1601–1833", *American Journal of Sociology*, **112** (1): 195–230.

Esch, A. (1985), "Überlieferungs-Chance und Überlieferungs-Zufall als methodisches Problem des Historikers", *Historische Zeitschrift*, **240** (3): 529–70.

Granovetter, M. (1978), "Threshold models of collective behavior", *American Journal of Sociology*, **83** (6): 1420–43.

Hannan, M.T. and G.R. Carroll (2000), *The Demography of Corporations and Industries*, Princeton, NJ: Princeton University Press.

Hedström, P., R. Sandell and C. Stern, (2000), "Mesolevel networks and the diffusion of social movements: The case of the Swedish Social Democratic Party", *American Journal of Sociology*, **106** (1): 145–72.

Hillmann, H. (2008), "Mediation in multiple networks: Elite mobilization before the English Civil War", *American Sociological Review*, **73** (June): 426–54.

Hillmann, H. (2021), *The Corsairs of Saint-Malo: Network Organization of a Merchant Elite Under the Ancien Régime*, New York: Columbia University Press.

Hillmann, H. and B.L. Aven (2011), "Fragmented networks and entrepreneurship in late imperial Russia", *American Journal of Sociology*, **117** (September): 484–538.

Leifer, E. (1992), "Denying the data: Learning from the accomplished sciences", *Sociological Forum*, **7** (2): 283–99.

Lena, J.C. and M.C. Pachucki (2013), "The sincerest form of flattery: Innovation, repetition, and status in an art movement", *Poetics*, **41** (3): 236–64.

Lusher, D., J. Koskinen and G. Robins (2013), *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*, Cambridge, UK: Cambridge University Press.

Lutter, M. (2015), "Do women suffer from network closure? The moderating effect of social capital on gender inequality in a project-based labor market, 1929 to 2010", *American Sociological Review*, **80** (2): 329–58.

Martin, J.L. (2017), *Thinking Through Methods: A Social Science Primer*, Chicago, IL: University of Chicago Press.

Moody, J. and J. Coleman (2015), "Clustering and cohesion in networks: Concepts and measures", pp. 906–12 in J.D. Wright (ed.), *International Encyclopedia of the Social and Behavioral Sciences*, Amsterdam: Elsevier.

Owen, T.C. (1992), *RUSCORP: A Database of Corporations in the Russian Empire, 1700–1914*, 3rd edition, Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Padgett, J.F. and C.K. Ansell (1993), "Robust action and the rise of the Medici, 1400–1434", *American Journal of Sociology*, **98** (6): 1259–319.

Padgett, J.F. and P.D. McLean (2006), "Organizational invention and elite transformation: The birth of partnership systems in Renaissance Florence", *American Journal of Sociology*, **111** (5): 1463–568.

Padgett, J.F. and W.W. Powell (2012), *The Emergence of Organizations and Markets*, Princeton, NJ: Princeton University Press.

Paris-Jallobert, P. (2004), *Anciens registres paroissiaux de Bretagne, d'après les travaux de l'abbé Paris-Jallobert. Ille-et-Vilaine, tome VII: Saint-Malo.* CD-ROM. Rennes: Sajef.

Quinn, J. (2017), *In Search of the Phoenicians*, Princeton, NJ: Princeton University Press.

Richardson, L. (2007). Beautiful soup documentation. Dosegljivo, https://www. crummy. com/software/BeautifulSoup/bs4/doc/. [Dostopano: 7. 7. 2018].

Rossman, G., N. Esparza and P. Bonacich (2010), "I'd like to thank the Academy, team spillovers, and network centrality", *American Sociological Review,* **75** (1): 31–51.
Salganik, M.J. (2019), *Bit by bit*: *Social research in the digital age*, Princeton, NJ: Princeton University Press.
Schelling, T.C. (2006), *Micromotives and Macrobehavior*, New York: Norton.
Stark, D. and B. Vedres (2006), "Social times of network spaces: Network sequences and foreign investment in Hungary", *American Journal of Sociology,* **111** (5): 1367–411.
Stovel, K., M. Savage and P. Bearman (1996), "Ascription into achievement: Models of career systems at Lloyds Bank, 1890–1970", *American Journal of Sociology,* **102** (2): 358–99.
Wei, M.Y. (2020), "The similarity network of motion pictures", *Management Science,* **66** (4): 1647–71.
White, H.C. (1970), *Chains of Opportunity: System Models of Mobility in Organizations*, Cambridge, MA: Harvard University Press.
White, H.C. (1981), "Where do markets come from?", *American Journal of Sociology,* **87** (3): 517–47.
Wickham, H. (2016), rvest: Easily Harvest (Scrape) Web Pages. R package version 0.3.2. https://CRAN.R-project.org/package=rvest
Zuckerman, E.W., T.-Y Kim, K. Ukanwa and J. Von Rittmann (2003), "Robust identities or nonentities? Typecasting in the feature-film labor market", *American Journal of Sociology,* **108** (5): 1018–74.

Katharina Burgdorf and Henning Hillmann - 9781789906851

19. Digital data and methods

Viktoria Spaiser

Digital data and methods are becoming increasingly ubiquitous. Almost all aspects of people's lives are now digitized: population registers are digital; health records are digital, criminal records are digital, employment and education registers are digital, all our interactions with authorities (whether on a local, regional or national level) are increasingly digital and hence produce digital records; most of our economic transactions (purchases, bank transfers, sales) are increasingly digital and hence produce digital records; our entertainment consumption is increasingly digital (Netflix, Spotify) and hence produce digital records and even our social life becomes increasingly digitized with social media like Facebook, Instagram, WhatsApp and so on. While there are still important realms that remain off the digital grid (e.g. elections in most countries, real-life social interactions and so on.), the vastness of activities that are now digitized can hardly be ignored. This creates both opportunities and challenges for social science researchers and sociologists in particular (Golder and Macy 2014, Hampton 2017). In this chapter I will discuss how analytical sociology can help to harness the opportunities and respond to the challenges and why and how analytical sociology should embrace digital data and methods in its repertoire of methodological approaches. I will first define and discuss what digital data and digital methods are, then I will discuss the opportunities and challenges and how analytical sociology can help to harness the first and deal with the latter and finally I will discuss what digital data and digital methods can offer to analytical sociology.

1. WHAT ARE DIGITAL DATA AND DIGITAL METHODS?

Digital data is, strictly speaking, any type of data that is digitized. The so-called big data can be subordinated under digital data, but it is not synonymous. Big data is often defined as "high volume, high velocity, and/or high variety [i.e. highly unstructured] information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization" (Laney 2001). Digital data however does not need to be of high volume; a small sample of digital data (e.g. a sample of Instagram images) can be of equal interest to researchers. Digital data is also not necessarily high velocity, i.e. real-time data, it can be equally historical digitized data. Finally, digital data is not necessarily high variety, i.e. unstructured, consisting of various data types. Digital data can be quite narrowly defined and structured within a specific research project, e.g. register data. But in its entirety digital data is displaying great variety. Such data can be:

- numerical (such as records from sensors),
- textual (such as social media posts or digitized archive records),
- images (such as pictures uploaded by users on Instagram),
- videos (such as videos uploaded by users on YouTube),
- spatial (such as latitude and longitude records),

Viktoria Spaiser - 9781789906851

- temporal (such as time records attached to social media posts),
- relational data (such as retweets, mention or response tracks),
- experimental data from (large-scale) online experiments (such as recorded choices, actions and interactions),
- emojis (for instance used on social media to express emotions non-verbally), and
- program code (for instance provided openly accessible through platforms such as GitHub).

All this data can be of great importance to research the social world. Sensor data, such as heart rate, sleep patterns and smartphone-based proximity measures may be of little relevance on their own, but they can be important measures when combined with other data. Sekara and Lehmann (2014) used, for instance, proximity sensor data to study the strength of friendship ties.

Textual data gives us invaluable insights about people's thoughts, attitudes, values, grievances, preferences, lifestyles, experiences and interactive discursive practices (Evans and Aceves 2016). Asher et al. (2019) studied, for instance, how people react to parliamentary discussions of popular e-petitions and how polarization can emerge in these conversations using tweets. Textual data can even be used to infer indirectly socio-demographic characteristics such as the age, gender and income of the social actors (Preoţiuc-Pietro et al. 2015; Hinds and Joinson 2018).

Through user-generated images we can also study people's views of the world, their perceptions, their experiences, and their self-representation. Often these images come with some textual data that add further information to the visual data. Zappavigna (2016), for example, used, Instagram images to study the representation of motherhood in public online spaces.

Video data combines visual and audio data and can provide extremely valuable information on events, including actors and their interactions; it can also provide insight into people's attitudes, perceptions and experiences as well as into discursive strategies, for instance, in political debates. English et al. (2011) studied, for instance, persuasion appeals in viral YouTube videos during the 2008 US presidential campaign.

Spatial data can provide an additional analytical feature that can be quite crucial for understanding social phenomena, since geographic contexts can contain crucial information, such as marking areas that are prosperous versus those that are disadvantaged. Social actors operate within these geographic spaces, and these geographic spaces have an impact on actions and how these actions aggregate. Gruenewald et al. (2006), for instance, used geo-spatial register data to demonstrate that alcohol outlets directly affected and moderated violence associated in disadvantaged neighbourhoods as well as in rural middle-class areas.

Similarly, temporal data provides important context information for social phenomena. Temporal data in particular is essential if it signifies crucial events, disruptions and changes that impact the social world; they can allow for natural experiments to be studied, where they divide data into before and after the natural experiment "intervention" subsets. Phan and Airoldi (2015) studied, for instance, friendship network formations and dynamics on Facebook with a natural disaster (hurricane) as a natural experiment "intervention" to show that social relationships may serve as a coping and resilience-building mechanism to deal with high-stress situations.

Relational data is in particular of interest to sociologists, since it records people's social behaviour, their interactions, their social connections and the network structure of our social

Viktoria Spaiser - 9781789906851

worlds. Williams et al. (2015), for instance, used Twitter data to establish a social network of Twitter users involved in the climate change debate and revealed the strong polarization that characterises the debate online.

Digital experimental data is generated through (large-scale) online experiments and is extremely valuable since it potentially allows to draw causal conclusions. The online experimental setups allow for scaling that is unachievable with conventional lab experiments (Kramer et al. 2014). The data usually records some form of actions, interactions and choices. Tsvetkova et al. (2016), for instance, created a web application and smartphone app that gamified Schelling's (1971) famous segregation model. Study participants had to move their virtual game piece on a lattice grid, considering the rules of the game and the actions of the other players. Using this application in several experimental studies with secondary school students in Sweden they generated unique digital experimental data that allowed novel conclusions on segregation dynamics to be drawn (Tsvetkova et al. 2016).

Emojis might not seem to be worth categorising as distinct data, but in fact they are. Emojis are neither genuine textual data nor genuine image data; they are digitally codified symbolic data that provide a whole layer of extra information about people's emotional states, attitudes, perceptions, preferences etc. Kaye et al. (2017), for instance, show how emojis can provide invaluable socio-psychological information about people and their online behaviour.

Finally, some might find it surprising that code as data was included, and that it is considered relevant for social science research. But, indeed, code can provide crucial insights into the algorithms that increasingly influence not only people's decision making and actions (e.g. algorithmic extremism),[1] but also institutional decision making and actions (e.g. predictive policing).[2] Algorithms are not neutral machine code; they are written by humans, and they serve certain purposes and have real-life consequences. The critical study of code is an emerging field that is certainly of great importance in our world that is increasingly dominated by human-machine interactions. Miller (2019), for instance, critically analysed an algorithm that allegedly predicted the likelihood of terrorist strikes. Generally, human-machine interaction deserves a greater focus in future social science research.

Digital methods are methods that allow us to process and analyse digital data. The specific method obviously depends on the type of data to be analysed; clearly textual data needs other approaches than relational data or image data. Computational social science is a whole new interdisciplinary field with a focus on computational digital methods that process and analyse, typically but not exclusively big data, i.e. observational large, unstructured digital data. It includes natural language processing and text mining as methods for processing and statistically analysing textual data, image processing for visual data (images and videos), network analysis for relational data and machine learning, which can be applied to various data types, including textual data, image data, video data, emojis data, audio data, sensor data and so on (Kreuschnigg et al. 2018). If the data is not observational but, for instance, generated through (large-scale) online experiments, then (conventional) statistical methods can be used to analyse the data.

---

[1] Algorithmic extremism is the online radicalisation of users that is driven by algorithmic recommendation systems (Ledwich and Zaitsev 2019).

[2] Predictive policing is the application of analytical techniques—particularly quantitative techniques—to identify likely targets for police intervention and prevent crime or to solve past crimes by making statistical predictions (Perry et al. 2013). Several predictive policing software is already in use across the world.

Viktoria Spaiser - 9781789906851

Machine learning is itself a large subfield and includes a great variety of methods and approaches, which can be extremely valuable for social researchers (Molina and Garip 2019). There are three main approaches in machine learning: supervised learning (e.g. random forests), unsupervised learning (e.g. topic modelling) and reinforcement learning (e.g. deep q-learning). Irrespective of the specific paradigm, the main idea is always that the computer program uses data to figure out the model to predict data rather than the human specifying the model. The difference in the approaches reflects the different ways for the machine to learn the model. The computer can learn by example (supervised learning), or by self-directed exploration of patterns (unsupervised learning), or by trial and error with cumulative rewards for correct estimations and punishment for error (reinforcement learning). These techniques are immensely powerful in uncovering hidden, often non-linear, patterns in the data and in making accurate predictions in certain application areas. But most of the time these methods are black boxes that do not reveal what the model is that the machine has learned, and this of course hinders understanding and interpretation. Research is ongoing in developing machine learning techniques that provide information about the machine-learned models (Molnar 2019, Murdoch et al. 2019).

Finally, digital methods can also include qualitative approaches, such as digital ethnography (Murthy 2008), where ethnographic methods, including participatory observations, are applied to online spaces, discourse analysis of digital textual data, qualitative visual methods such as visual discourse analysis (Wang 2014) of digital image data and so on.

2. ANALYTICAL SOCIOLOGY AND DIGITAL DATA'S OPPORTUNITIES AND CHALLENGES

Digital data are increasingly gathered by cheap and numerous information-sensing web services and digital devices. They are often quasi-automatic side-products of any use of digital technology. But, nevertheless, that data provides great opportunities for social science researchers as the data is often rich and fine-grained (e.g. social interactions continuous in time), providing simultaneous information about what, where, who, when and with whom, e.g. what has been discussed (topics) and in what way (factual, emotional, polarized), who has been involved in the debate, when the debate took place, where the actors are located who participated in the debate and whom the discussion participants interacted with in what way.

Digital data moreover provides records of genuine, real-life, often real-time events (e.g. political protests or elections), processes (e.g. economic transactions), human behaviour and expressions of attitudes, values, emotions and knowledge, since it is usually not derived from standardised surveys or from interviews that nudge people to engage with certain topics that the researcher finds interesting. Hence the data allows us to understand and monitor what is going on in real time and in real life.

As such, this data is of outstanding value when describing the social world, at least as far as it is represented virtually. Manzo (2014) has suggested seven principles to guide analytical sociology research, and the second principle is hereby the fundamental one. It requires empirical evidence and hence data input to describe the phenomenon to be studied as precisely and completely as possible. This step follows the first step where the main concepts are defined and hypotheses are formulated. Given the usually observational nature of the digital data, it is often of limited use for testing hypotheses and theories; its greatest value

Viktoria Spaiser - 9781789906851

lies in description and, specifically, in using digital methods to find often hidden patterns in the data that give us descriptively a deeper insight into the social world dynamics. While descriptive analyses are often dismissed, analytical sociology argues that descriptive analysis is essential and its contribution to good sociological research should be acknowledged and valued (Bearman 2012; Manzo 2014). Analytical sociology can, therefore, help social science researchers to harness the potential of digital data by first facilitating the understanding of the great value of the data, even if it is observational data.

The descriptive value of the data goes beyond describing univariate attributes, characteristics and dynamics; it extends to identifying relations, both bivariate and multivariate ones. From an analytical sociology perspective these identified relations cannot be used to establish causal links and explanations, but they are nevertheless an important first step in any analysis that aims to specify the mechanisms that generate certain macro-level patterns from micro-level activities (Manzo 2014). Observational data is often dismissed as it does not allow for establishing causal relations. Data-driven approaches produce relations and patterns without explanations. However, if researchers adopt the analytical sociology perspective, then explanation is not sought at this second stage of the research process. From this analytical sociology perspective observational data such as digital data is not of lesser value than survey data or other data that social scientists often collect. On the contrary, the invaluable potential of digital data to discover bivariate and multivariate relations; using methods such as machine learning as one step in an analytical sociology research process, is much more evident once it is accepted that finding causality is not the goal at this stage of research. The data is then a rich well for explorative analyses to make discoveries of potentially unexpected patterns, which can help to develop generative models that specify the mechanisms at work, which can be then tested at later stages of the research process.

Digital data, however, confronts the research with great challenges. There are technical challenges, such as the storage and computational power required for the vast volume of digital data that often require technical solutions, but there are also challenges that are of a more fundamental nature. The unprecedented volume of data and the often observational nature of the data often calls for data-driven approaches that are frequently black-box methods, such as machine learning, where it is not clear what patterns the machine exactly learns to arrive at certain predictions.

It has been repeatedly suggested that computational social science in particular needs to be grounded more in theory rather than just pure data to move beyond descriptions and predictions (Kreuschnigg et al. 2018). Most researchers across disciplines working with digital data disagree with Anderson's (2008) declaration of the end of theory when he remarked: "Petabytes allow us to say: 'Correlation is enough'. We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot." And yet a great share of computational social science is still done in this fashion. The research process is essentially stopped at the description stage. This is partly the result of computer scientists working in the field not being familiar with social theory, and partly because there is no unifying theory that computational social science research could build upon.

Analytical sociology could provide such a unifying meta-theory that can be applied across a range of social phenomena studied by computational social scientists, guiding the research by allowing research questions to be formulated, identifying the data features that are of interest

Viktoria Spaiser - 9781789906851

to answer the research questions and by putting the data analysis within a greater research process, guided by the seven principles outlined by Manzo (2014). What makes analytical sociology particularly appealing to computational social scientists is the fact that both rest implicitly on structural individualism, where individual actions result in macro-level patterns inherent in social structures and macro-level dynamics, but at the same they are constrained and framed by these social structures (Boudon 1986, Hedström and Bearman 2009, Kreuschnigg et al. 2018). Adopting analytical sociology as a meta-theory would allow for research that includes digital data to be research-question driven rather than data-driven.

Moreover, if the data analysis is not the sole centre piece of research, but one essential stage within a more comprehensive research process, as outlined by Manzo (2014), then it is also easier to acknowledge the limitations of data-driven approaches due to limitations of digital data, or, indeed, any data. Digital data often requires special precaution and critical reflection though, because it is often not fully transparent (Fuchs 2017). For instance, if a researcher collects Twitter data, they rely on Twitter for access to the data. But the social media company will apply quality filter and rate limitations to the data it provides access to for free, and these filters are not fully transparent. The implicit problematic issue with many digital datasets is ultimately data ownership (Ruppert 2015). Increasingly the data is owned by large social media companies and researchers are at their mercy for access to the data (Felt 2016). This can create an imbalance where research with digital data is increasingly conducted by social media companies at their terms and with their agendas, often not meeting academic sector ethical research standards. An example for such a study is the much criticised large-scale experiment on emotional contagion conducted with Facebook users without their consent (Kramer et al. 2014). I will come back to this point in my concluding remarks.

Beyond these issues with respect to data ownership, accessibility and transparency, data, any data, is never complete and will always include errors or irrelevant information (noise) and hence data cannot be seen as something absolute, pure and equivalent to the objective truth. Researchers using, for instance, Twitter data must acknowledge that the Twitter user space is not representative of any population and includes numerous bot accounts, which obviously can be misleading if mistaken for humans. Finally, data is always a snapshot and record of the world as it is, not as it could be if certain mechanisms and parameters were to be changed slightly. Mechanism-driven approaches as suggested by analytical sociology would allow us not only to see the world as it is, but explore what the world could be if we were to introduce certain changes based on a rigorous understanding of the mechanisms at work rather than some potentially spurious correlations.

3. DIGITAL DATA AND METHODS FOR ANALYTICAL SOCIOLOGY

Thus far, I have discussed how analytical sociology can help to harness the opportunities and deal with some of the challenges of digital data and methods. But, equally, analytical sociology can greatly benefit from the available and still emerging digital data and methods. Manzo (2014) argues in his seven principles of analytical sociology that analytical sociology ought to "mobilise the best quantitative and qualitative empirical information available and use the technical tools best suited to describing the facts to be explained" (p. 7). Arguably, digital data and methods will be the best available empirical data and methods for a range of phenomena of interest. As has been discussed in previous chapters, analytical sociology is interested in

Viktoria Spaiser - 9781789906851

finding and understanding usually individual-level mechanisms in order to explain higher-level emergent phenomena (Hedström 2005). From the previous elaborations it should be clear that digital data and methods can at the very least greatly contribute to the second principle or stage of analytical sociology research, namely providing fine-grained and accurate descriptions of the phenomena to be explained.

This is clearly valuable on its own, but digital data and methods can do more. I will argue that digital data and methods can also contribute to the third and quite crucial stage or principle of analytical sociology research, namely "to formulate a 'generative model', that is a model of a (set of) mechanism(s), where a mechanism is a set of entities and activities likely to trigger a sequence of event (i.e. a process) likely to bring about the outcomes(s)" (Manzo 2014, p. 7). Furthermore, it can be of great value for principle four, which requires "a realistic description of the relevant micro-level entities (P4a) and activities (P4b) assumed to be at work as well as of the structural interdependencies (P4c) in which these entities are embedded and their activities unfold" in order to formulate a generative model (Manzo 2014, p. 8). Finally, digital data can be of great benefit for principles six and seven. Based on principle six, digital data can be used to validate an agent-based model that captures the mechanisms specified in the generative model. Ideally the computational model should reproduce the empirical facts described by the digital data. Following principle seven, digital data can be used to empirically ground the agent-based simulation model (digital data can be injected into the simulation model) to make sure its parameter space is resembling the real world as much as is reasonable and possible and then to analyse how the real-life data parameter setting impacts the model behaviour and outcomes (ibid.).

In fact, one of the great advantages of digital data in comparison to more conventional data such as survey data is that, if it is actually large, it can be easily split in various subsets, e.g. one subset to calibrate the individual-level parameter and behaviour rules of the agent-based model and one subset to test the predicted macroscopic consequence of the agent-based model calibrated with the first subset of the data. Any single survey-based dataset is, by contrast, often not big or diverse enough to allow for these different tasks to be performed with the same dataset.

I will now use a computational social science study to demonstrate more explicitly how digital data can be used at various stages of the research process and how it can make a contribution to various principles of analytical sociology research. The study by Romenskyy et al. (2018) researched the political polarization and crisis in Ukraine in 2014 that ultimately resulted in a militarised conflict in separatist territories of Ukraine, using Twitter data and a simple agent-based model for opinion dynamics.

In the study the authors first defined and specified the theoretical concepts used in studies of opinion polarization, thereby following the first principle of analytical sociology. They then conducted descriptive analyses of the opinion polarization in Ukraine as they presented within the Twittersphere, hence following the second stage and principle of analytical sociology. The researchers used Twitter users' tweets to classify and position the users within a continuous two-dimensional opinion space, composed of opinions toward the West/Europe on one dimension and opinions toward the East/Russia on the other dimension. They did this classification repeatedly over time to detect changes in opinions over time. They also explored the geo-spatial distribution of the opinions. This descriptive analysis established and quantified the phenomena to be explained, namely the increasing opinion polarization over time.

Viktoria Spaiser - 9781789906851

According to Manzo (2014) the next stage, following the third principle of analytical sociology, ought to be the formulation of a generative model, specifying the mechanisms that lead to opinion polarization. Indeed, the authors developed the so-called bounded confidence XY model of opinion polarization (Romenskyy et al. 2018), a simplified agent-based model where each agent is represented through a spinning vector placed on a two-dimensional lattice of finite size. The agents interact with their neighbours and consequently update their opinion direction (polar angle) following three main mechanisms based on a set of opinion polarization theories:

1. Social interactions are introduced via local mean field, i.e. the new orientation of each vector is calculated as a mean of the average direction of the neighbours and agent's own orientation (social comparison theory)
2. Noise is added to model the level of conformity of the individual, where zero noise corresponds to full conformity and noise greater zero allows an individual to deviate from the group opinion (biased assimilation theory)
3. Interaction is selective so that the vectors align only with those neighbours whose orientation (opinion) deviates by an angle less than some fixed value from their own opinion vector (homophily theory).

But, what is interesting in this study is that a fourth mechanism was introduced to the model, not based on an established theory, but inspired by the explorative descriptive analysis of the Twitter data. Specifically, when conducting sentiment analysis with the Twitter data as part of the political classification analysis, they discovered that the emotionality of tweets jumped suddenly up in February 2014, around the time when the protest escalated after 100 protesters were shot by snipers on a single day. After this escalation the emotionality of the tweets remained on a high level and never reached the level before February 2014 again. Noting that other research has suggested that emotions may play an important role in opinion polarization, they decided to include emotionality as a parameter in the agent-based model. In particular, the fourth mechanism implemented in the model weighted the contribution of each interacting agent by the emotional intensity of its opinion. Emotions hence serve as an amplifier in the model. Here, digital data helped to formulate a generative model because the explorative analysis of these data suggested that another mechanism, not specified within existing social theory, should be considered.

Following the fourth principle (Manzo 2014), the authors provide in their paper (Romenskyy et al. 2018) a precise description of the micro-level entities and their activities as represented in the Twitter data, as well as the structural interdependencies within which the social processes unfold. The structural interdependencies that are in the focus within this study include the geographical division in Ukraine, where the North and West are inhabited predominantly by Ukrainian-speaking citizens who exhibit stronger political alignment with the West and Europe, while the South and East is inhabited by large groups of Russian-speaking citizens who feel a stronger affiliation with Russia. This structural interdependency is also captured in the agent-based model through boundary conditions, set by two fixed rows of agents on each side. The boundary vectors on the left and upper rows are permanently oriented to the left, imitating a bias toward "West," and the right and bottom rows are fixed in the opposite direction, mimicking a bias toward "East." This boundary condition hence introduces a spatially dependent local bias that can account for geographical inhomogeneity in opinions

Viktoria Spaiser - 9781789906851

we see empirically in Ukraine. Digital data, i.e. Twitter data, was used here to make sure that the empirical geographical division in Ukraine can also be established from the Twitter data, hence validating the Twitter data and at the same time describing the data taking structural independencies into account.

The fifth principle of analytical sociology has also been implemented in this study as the authors translated the generative model into an agent-based computational model (Romenskyy et al. 2018). But, as suggested by principles six and seven, the authors also use Twitter data to validate their computational model (principle six) and to parametrize the computational model (principle seven). Specifically, they used a random subsample of weekly distributions of the overall users' emotional intensity score to parametrise the emotionality parameter of the agent-based model. In each simulation, emotionality values were assigned to agents at random, according to the obtained discrete weekly distributions of overall emotional intensity, and kept constant throughout the simulation. One could argue that the authors could have attempted to use the data as far as possible to also set the other parameters (i.e. noise, homophily). However, the data that Twitter provides access to for free is only a random 1% of all public tweets and hence it is incomplete and does not necessarily allow the tracing of individual Twitter users over time. This, however, would be required to establish the empirical values for the noise model parameters, i.e. how many Twitter users are willing to change their opinion. The extent of homophily, on the other hand, could have been established approximately by investigating, for instance, retweet networks. For validation purposes the outcomes of the simulation, i.e. distribution of opinions and their spatial clustering, was compared to the empirical evidence established through the subset of Twitter data withheld for model validation purposes. So here digital data was again of great importance to assess to what extent the computational agent-based model was suitable to explain the macro-level patterns, i.e. opinion polarization, we can observe. Although the model can be deemed quite simple, it predicts macro-level patterns surprisingly well.

The research in general revealed the underestimated importance of emotions and their intensity levels in opinion polarization processes. The authors write:

> We demonstrated that the tendency of territorial splitting is conditioned by the high conformity and homophily in the society and is driven by the growth in emotional intensity […]. Specifically, while our analyses confirm the importance of social comparison, homophily, persuasive argument and biased assimilation mechanisms and their specific interactions for polarization, they also show that these mechanisms are not sufficient to ignite the extreme societal polarization we can observe, for instance, in Ukraine. The emotional intensity is a key reinforcement mechanism that has to be added. (p. 10).

Throughout the analysis process digital data played a key role in establishing the facts, in developing the generative model and in implementing and testing the generative model. The key study result about the importance of emotional intensity was inspired by initial exploratory data analysis, which suggested the inclusion of the emotional intensity mechanism in the generative model in the first place. So (digital) data can play an immensely important role in a research approach informed by analytical sociology principles.

The study clearly has some limitations. One could, for instance, question to what extent Twitter data is presenting an accurate picture of the opinion polarization in Ukraine. I have already noted earlier that Twitter data is never representative of any given population. However, what are the alternatives? Survey data that captures representative opinions of a nation is only

Viktoria Spaiser - 9781789906851

sporadically available and not suitable to track fine-grained opinion dynamics over time and space. Hence, in this case, digital social media data is the best available data to study the given phenomenon. One could argue, however, that the authors could have considered other social media, such as Odnoklassniki, which is a very popular social media site in Russia and former Soviet Union republics. In an expanded study the authors could have gathered data from that alternative platform and conducted similar analyses to see whether they could replicate the findings obtained with Twitter data. One could also criticise the generative model as too simplistic to capture real-life dynamics accurately, but it should be acknowledged that this very simple model replicated the empirical evidence extremely well and generated novel, important insights. In the next step more complex models and mechanisms can be explored.

In any case, this study is a very good example of how computational social science can implement analytical sociology principles to generate insights that go beyond description and prediction and facilitate actual understanding. It also shows how the full set of principles of analytical sociology research can be realised within a single study that involves digital data and methods. And, given the role the data played throughout the research process, including in the development of the generative model, this study demonstrates the great value that digital data and methods can have for analytical sociology.

4. CONCLUDING REMARKS

I have explained in this chapter what digital data and digital methods are and discussed the opportunities and challenges that they pose for social science scholars and how analytical sociology can help to embrace the opportunities and respond to some of the challenges. Using Manzo's (2014) seven principles of analytical sociology, I have also explained how at different stages and in response to different principles digital data and methods can contribute to social research based on analytical sociology. An exemplary study by Romenskyy et al. (2018) was used to demonstrate the great value of digital data and methods for analytical sociology approaches following the seven principles developed by Manzo (2014).

Digital data and methods are still rapidly developing, and they will certainly become even more central for any social science research in the near future. For instance, it is becoming increasingly evident that smart mobile devices, such as smartphones, and in the near future, potentially, also smart glasses, watches, wearables etc. will be increasingly used to collect data, and not only observational data but also field-experimental data (Spaiser et al. 2019). Similarly, (large-scale) online experiments (Centola 2010; Tsvetkova et al. 2016) are likely to become more prominent in the near future. This will open up new avenues for exciting sociological research and, in particular, for mechanism-driven research that analytical sociology is promoting, since (causal) mechanisms on the individual level can be best empirically studied using (field) experiments. Hence, these experimental studies will allow us to formulate better, more accurate and realistic generative models as they will boost the third and fourth principles of analytical sociology.

But, observational data, such as social media data, is likely to remain highly relevant in future social research. Furthermore, traditional social science data collection such as surveys are increasingly informed by and merged with digital data (Couper 2017), and this is likely to accelerate in the coming years. For instance, online surveys are becoming increasingly common; this includes surveys with social media users, which allows the merging of

Viktoria Spaiser - 9781789906851

survey data and social media data (Wojcik and Hughes 2019) harnessing the benefits of both data collection forms.

The potential of observational data such as social media data for analytical sociology has been demonstrated here, but issues remain, and they go beyond methodological or theoretical questions. The most critical issue is that these data are proprietary data and as such not easily available to social scientists unless they have a sufficiently large budget to purchase the data. There is a reason Twitter data features so prominently in social science research; it is because it is the most easily accessible social media data. Other large social media companies, such as Facebook, have extremely limited access to their data, which has made research with the data increasingly difficult if not impossible, unless one has some privileged access through collaborations with Facebook (King and Persily 2020). This has happened not least because of data-related scandals such as the Cambridge Analytica scandal that also involved researchers (Lapowsky 2019). But it should be acknowledged that data is what generates profits in these companies and hence it should be of little surprise that these companies limit free access to the data. In fact, Twitter is increasingly restricting free access to its data as well, as do other social media companies. What does that mean for future social science research? Most likely researchers will increasingly have to obtain funds that will allow them to purchase certain datasets. Also, we may see more collaborations between researchers and social media companies that will provide privileged data access to selected researchers. It remains to be seen how this will affect the direction of social science research in the future. Critical reflexivity will be crucial to maintain independence of social science research, including analytical sociology research that involves proprietary digital data.

REFERENCES

Anderson, Chris (2008), "The end of theory: The data deluge makes the scientific method obsolete". *Wired*, June 2008, https://www.wired.com/2008/06/the-end-of-theo/

Asher, Molly, Christina Leston-Bandeira and Viktoria and Spaiser (2019), "Do parliamentary debates of e-petitions enhance public engagement with parliament? An analysis of Twitter conversations", *Policy & Internet*, **11** (2): 149–71.

Bearman, Peter (2012), "On analytical sociology", *Sociologica*, 1, https://doi.org/10.2383/36895

Boudon, Raymond (1986), *Theories of Social Changes: A Critical Appraisal*, Cambridge and Oxford: Polity Press/ Basil Blackwell.

Centola, Damon (2010), "The spread of behavior in an online social network experiment", *Science*, **329** (5996): 1194–7.

Couper, Mick P. (2017), "New developments in survey data collection", *Annual Review of Sociology*, **43**: 121–45.

English, Kristin, Kaye D. Sweeter and Monica Ancu (2011), "YouTube-ification of Political Talk: An Examination of Persuasion Appeals in Viral Video", *American Behavioural Scientist*, **55** (6): 733–48.

Evans, James A. and Pedro Aceves (2016), "Machine translation: Mining text for social theory", *Annual Review of Sociology*, **42**: 21–50.

Felt, Mylynn (2016), "Social media and the social sciences: How researchers employ big data analytics", *Big Data & Society*, **3** (1): 1–15.

Fuchs, Christian (2017), *Social Media. A Critical Introduction*, 2nd edition, London: SAGE.

Golder, Scott A. and Michael, W. Macy (2014), "Digital footprints: Opportunities and challenges for online social research", *Annual Review of Sociology*, **40**: 129–52.

Gruenewald, Paul J., Bridget Freisthler, Lillian Remer, Elizabeth A. LaScala and Andrew Treno, (2006), "Ecological models of alcohol outlets and violent assaults: Crime potentials and geospatial analysis, *Addiction*, **101** (5): 666–77.

Hampton, Keith N. (2017), "Studying the digital: Directions and challenges for digital methods", *Annual Review of Sociology*, **43**: 167–88.

Hedström, Peter (2005), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge: Cambridge University Press.

Viktoria Spaiser - 9781789906851

Hedström, Peter and Peter Bearman (2009), "What is analytical sociology all about?", in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press, pp. 3–24.
Hinds, Joanne and Adam N. Joinson (2018), "What demographic attributes do our digital footprints reveal? A systematic review", *PLoS ONE*, **13** (11): e0207112.
Kaye, Linda K., Stephanie A. Malone and Helen J. Wall (2017), "Emojis: Insights, affordances, and possibilities for psychological science", *Trends in Cognitive Sciences*, **21** (2): 66–8.
King, Gary and Nathaniel Persily (2020), "Unprecedented Facebook URLs dataset now available for academic research through Social Science One", *Social Science One*, https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one
Kramer, Adam D.I., Jamie E. Guillory and Jeffrey T. Hancock (2014), "Experimental evidence of massive-scale emotional contagion through social networks", *PNAS*, **111** (24): 8788–90.
Kreuschnigg, Marc, Niclas Lovsjö and Peter Hedström (2018), "Analytical sociology and computational social science", *Journal of Computational Social Science*, **1**: 3–14.
Laney, Doug (2001), *3D Data Management: Controlling Data Volume, Velocity, and Variety*, Application Delivery Strategies, META Group Inc., File 949.
Lapowsky, Issie (2019), "How Cambridge Analytica Sparked the Great Privacy Awakening", *Wired*, https://www.wired.com/story/cambridge-analytica-facebook-privacy-awakening/
Ledwich, Mark and Anna Zaitsev (2019), "Algorithmic extremism: Examining YouTube's rabbit hole of Radicalization", arXiv: 1912.11211.
Manzo, Gianluca (2014), "Data, generative models, and mechanisms: More on the principles of analytical sociology", in Gianluca Manzo (ed.), *Analytical Sociology*: *Actions and Networks*, Chichester, UK: Wiley, pp. 4–52.
Miller, Flagg (2019), "Terrorist violence and the enrolment of psychology in predicting Muslim extremism: Critical terrorism studies meets critical algorithm studies", *Critical Studies on Terrorism*, **12** (2): 185–209.
Molina, Mario and Filiz Garip (2019), "Machine learning for sociology", *Annual Review of Sociology*, **45**: 27–45.
Molnar, Christopher (2019), *Interpretable Machine Learning*, GitHub, https://christophm.github.io/interpretable-ml-book/
Murdoch, W. James, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl and Bin Yu (2019), "Definitions, methods, and applications in interpretable machine learning", *PNAS*, **116** (44): 22071–80.
Murthy, Dhiraj (2008), "Digital ethnography: An examination of the use of new technologies for social research", *Sociology*, **42** (5): 837–55.
Perry, Walter, L., Brian McInnis, Carter C. Price, Susan C. Smith and John S. Hollywood (2013), *Predictive Policing. The Role of Crime Forecasting in Law Enforcement Operations*, RAND Corporation.
Phan, Tuan Q. and Edoardo M. Airoldi (2015), "A natural experiment of social network formation and dynamics", *PNAS*, **112** (21): 6595–600.
Preoţiuc-Pietro, Daniel, Svitlana Volkova, Vasileios Lampos, Yoram Bachrach and Nikolaus Aletras (2015), "Studying user income through language, behaviour and affect in social media", *PLoS ONE*, **10** (9): e0138717.
Romenskyy, Maksym, Viktoria Spaiser, Vladimir Lobaskin and Thomas Ihle (2018), "Polarized Ukraine 2014: Opinion and territorial split demonstrated with the bounded confidence XY model, parametrized by Twitter data", *Royal Society Open Science*, **5** (8): 171935.
Ruppert, Evelyn S. (2015), "Who owns Big Data?", *Discover Society*, **23**. https://discoversociety.org/2015/07/30/who-owns-big-data/
Schelling, Thomas C. (1971), "Dynamic models of segregation", *Journal of Mathematical Sociology*, **1**: 143–86.
Sekara, Verdran and Sune Lehmann (2014), "The strength of friendship ties in proximity sensor data", *PLoS ONE*, **9** (7): e100915.
Spaiser, Viktoria, Davide Luzzatti and Aliki Gregorious (2019), "Advancing sustainability: Using smartphones to study environmental behavior in a field-experimental setup", *Data Science. Methods, Infrastructure and Applications*, **2** (1–2): 277–90.
Tsvetkova, Milena, Olof Nilsson, Camilla Öhman, Lovisa Sumpter and J.T David (2016), "An experimental study of segregation mechanisms", *EPJ Data Science*, **5** (1): 4.
Wang, Jiayu (2014), "Criticising images: Critical discourse analysis of visual semiosis in picture news", *Critical Arts*, **28** (2): 264–86.
Williams, Hywel T.P., James R. McMurray, Tim Kurz and F. Hugo Lambert (2015), "Network analysis reveals open forums and echo chambers in social media discussions of climate change", *Global Environmental Change*, **32**: 126–38.
Wojcik, Stefan and Adam Hughes (2019), "Sizing up Twitter users," Pew Research Center, https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/
Zappavigna, Michele (2016), "Social media photography: Construing subjectivity in Instagram images", *Visual Communication*, **15** (3): 271–92.

Viktoria Spaiser - 9781789906851

20. Advances in mediation analysis

Kinga Makovi and Christopher Winship

1. INTRODUCTION

The extensive causal inference revolution in statistics, epidemiology and other social science fields based on Rubin's potential outcome model and Pearl's directed acyclic graphs (DAGs) has not as yet penetrated the research of analytical sociologists (for a notable exception, see Manzo 2022). We believe that this is a mistake. Although the models considered within this causal inference literature are quite often theoretically simplistic, if not naive, what has been demonstrated is that even in the most basic situations it can be difficult, if not impossible, to strongly warrant a claim of a causal effect. This is true in case of treatments, as is common in this literature, or mediating mechanisms, which are of interest to analytical sociologists.

The title of this chapter could be "Dissecting social data: Advances in mediation analysis," deliberately referencing Peter Hedström's (2005) foundational book for analytical sociology, *Dissecting the Social: On the Principles of Analytical Sociology*. Whereas Hedström's book focuses on the theoretical dissection of the social, our purpose in this chapter is to consider how to "dissect" social data. Obviously, social science and sociology in particular are both about theory development and empirical analysis.

While this distinction is made often, we argue that theory and empirics is a false dichotomy: the data we collect and analyze are clearly informed by our theories of the causal structure of the world, and our empirical investigations in turn inform our thinking about the world's causal structure. The key question we address is *when* is it possible to assert, even in fairly simple situations, that empirical data in some context support the existence of a particular, hypothesized causal mechanism. This is not just a matter of showing that data are consistent with the presence of the mechanism but, just as importantly, demonstrating that the data are inconsistent with other plausible explanations.

We discuss two papers that have won the Analytical Sociology Association's prestigious Merton Award. We have chosen these papers since, as is true of much research within this subdiscipline, they are concerned with mechanisms. We demonstrate that the mechanisms in question could be conceived of as mediating variables. Although analytical sociology has argued for more complex theoretical conception of what is meant by a social mechanism, something that we applaud, empirical mediation analysis is still an important component of this sociological approach. We use these award-winning papers as evidence that analytical sociologists need to understand the implications of recent research on mediators for their own scholarship.[1]

An enormous amount of work on mediation analysis has been done in the past two decades using the potential outcome model and directed-acyclic graphs (DAGs). The resulting body of work has two opposing upshots—a glass half-full and a glass half-empty view. On the positive

[1] See Chapter 10 of Morgan and Winship (2015) for an extended discussion of the debate between analytical sociologists and others on how we think about causal mechanisms.

King Makovi and Christopher Winship - 9781789906851

side, the front door criterion and its extensions (Pearl, 2009; Knight and Winship, 2013) demonstrate that elaborating on mechanisms in fact helps rather than hinders the identification of causal effects. On the negative side, Pearl (2009), and Imai and collaborators (Imai et al., 2010a, b, 2011) establish that strong, typically unrecognized, theoretical assumptions underlie traditional methods of mediation analysis. Needed assumptions are not testable, and in many circumstances unrealistic. Fortunately, the literature has constructed methods for sensitivity analysis to determine how robust findings are to these assumptions, a major step forward. In addition, this line of work has provided new methods of analysis, in particular nonparametric approaches that make weaker assumptions. Thus, the literature has simultaneously opened new avenues to research, while also clearly showing the limitations, if not the fatal properties, of older methods.

We structure this chapter as follows. We start by introducing the empirical examples, primarily relying on the Merton Award–winning papers, but also add an extended empirical example of our own. We weave these three examples through the chapter to accompany key concepts as we introduce them. We then provide a brief introduction to the potential outcome model and DAGs. Morgan and Winship (2015) provide more extended discussions of each. Subsequently, we discuss different strategies for mechanism-based identification of causal effects. Finally, we make use of the extended empirical example to expose recent methodological developments in mediation analysis, arguably for the enrichment of analytical sociologists' methodological tool kit.

2. MECHANISMS: FROM STRATEGIC VEILING TO RISK COMPENSATION

The first example we consider is by Aksoy and Gambetta (2016). They start from a puzzling anecdotal observation, namely that Muslim women defy classical sociological theories of expressions of religiosity, which hold that the younger, the more educated, in urban settings and with high levels of exposure to less-religious others, limit their expressions of religiosity. They argue that such modernizing forces do not have a uniform effect on all Muslim women. They find that while those of average religiosity behave as predicted by the classical theories and veil less, and less strictly, women who are highly religious veil more, and more strictly, especially so when they are younger, single, more educated and live in urban settings and in communities with a less-religious majority.

The authors assert that this is the result of two strategies, or mediating mechanisms, the first of which is self-directed and the second is other-directed. They suggest that women use veiling as a commitment strategy to religious rules by either reducing their own temptation or their risk of exposure to the advances of others, which would both result in breaking religious rules of conduct and propriety—being the self-directed mechanism. They also suggest that women aim to signal to others, e.g., potential husbands or other, similarly minded families in their community, that they have remained pious and honourable in the face of modernizing forces—being the other-directed mechanism. In short, Aksoy and Gambetta theorize that effects of modernization and religiosity interact in such a way that more religious women in the modern world are more likely to adopt strategies that result in their veiling.

Second, we consider a paper by Cowan, who argues that the passing on of secrets depends on the sender's view on how the secret would be received (2014). Therefore, those who are anticipated not to be charitable toward (stigmatized) secrets will not hear them as a result of

King Makovi and Christopher Winship - 9781789906851

strategic information management. Advancing this argument, Cowan attempts a causal statement and asserts that it is selective disclosure that drives what information people have, and that this phenomenon is particularly important to understand how stigmatized information, or secrets, diffuse on social networks.

Motivations for sharing, or not, depend on two counteracting forces: one's need for support and shared self-understandings—which is probably associated with how close one is to a confidant—and one's strategic expectations about one's confidant's acceptance of the secret. In sum, (stigmatized) information sharing is mediated by one's confidant's acceptance of it, one's closeness to the confidant, as well as the interaction of acceptance and closeness.

Last but not least, we consider the unintended consequences of taking pre-exposure prophylaxis (commonly known as PrEP) on one's risk of contracting sexually transmitted diseases, such as chlamydia, gonorrhoea or syphilis. PrEP is a drug that significantly lowers one's chances of contracting HIV when exposed to it, and has been commercially available from 2012 in the US, and more recently elsewhere (Grant et al., 2010, 2014). In Western countries, PrEP has typically been prescribed to men who have sex with men (MSM), as one of the largest at-risk groups in terms of HIV exposure.

PrEP is the latest in a series of HIV-related interventions that has made headlines with claims of risk compensation, namely, that those who take PrEP might change their sexual behaviors and engage in higher-risk sexual practices, such as having more sexual partners while using protection less, mediating the relationship between PrEP status and sexually transmitted infection (STI) risk other than HIV, and consequently, diagnosis. These behaviors would change the diffusion of STIs and potentially alter the structure of sexual networks.

In all of these examples, the mediating mechanisms elaborate on "how" causal effects are brought about. From the perspective of analytical sociologists this is sociology at its best, creating a deep understanding of the causal chain over "just so" explanations, i.e., the achievement of theoretical depth. Note further that mediating mechanisms allow researchers to "walk the Coleman boat" to its fullest extent (Bearman, 2012; Ylikoski, 2017) and make connections between micro-level mechanisms (signalling, selective disclosure, risk compensation) to macro-level outcomes (the distribution of expressions of religiosity, awareness of (stigmatizing) secrets and sexual-health outcomes in the population). From the perspective of those who pursue mediation analyses, mediating mechanisms—and their formalization in the potential outcome model and their depiction as DAGs—allow for better understanding of the assumptions needed for identification, and rigorous analysis of whether the asserted causal claims are indeed warranted.

3. THE POTENTIAL OUTCOME MODEL

In what many call the revolution in causal inference, Rubin's potential outcome model and Pearl's DAGs have played central, but different, roles. The potential outcome model has been influential as it provides a rigorous definition for what it means for some variable, often called the treatment, to have a causal effect on an outcome. This has provided important insights into when the results of traditional empirical estimators, particularly those produced by regression, can or cannot be interpreted as causal effects.

We do not believe that the potential outcome model with its focus on manipulability at the level of the unit of analysis (often the individual) is the definitive way to think about causality (Russo, 2009; Runhardt, 2015; Rohlfing and Zuber, 2019). Instead, we are sympathetic to

King Makovi and Christopher Winship - 9781789906851

analytical sociology's commitment to a more sophisticated and pluralistic set of concepts of causality. What is important about the potential outcome model, however, is that it provides a rigorous definition of causality that can be useful for understanding the inferential properties of different methodological practices.

The potential outcome model, also called the counterfactual causal model, goes back to the 1920s (Neyman, [1923] 1990; Neyman et al., 1935), but it has been developed most completely by Don Rubin (1974, 1977, 1978, 1980, 1981, 1986, 1990). Conceptually, the basic approach is to extend the logic and language of experiments to non-experimental settings. At its core is the assumption that there is an outcome for every possible level of the treatment for any unit (most often individual), and that the treatment's causal effect may be theoretically defined based on differences in these potential outcomes, which correspond to different levels of the treatment.

Without loss of generality, we assume that there are two treatment statuses, treatment and control. More formally, let the variable T_i be defined as:

$$T_i = \begin{cases} 1 & \text{if individual } i \text{ is in the treatment condition,} \\ 0 & \text{if individual } i \text{ is in the control condition.} \end{cases}$$

Then[2]

$Y_i(1)$ is the outcome when i is in the treatment condition, and

$Y_i(0)$ is the outcome when i is in the control condition.

The individual causal effect (ICE) is:

$$\delta_i = Y_i(1) - Y_i(0).$$

This is the core quantity in theory. By relying on this quantity one can then define the average causal effect, that is, the average of individual causal effects over the population. Often this is the quantity that social scientists are most interested in, namely the impact of an intervention, policy or change over the population on some outcome.[3] More formally, the average causal effect (ACE), often called the average treatment effect (ATE), is:

$$\text{ACE} = \bar{\delta} = \sum_{i=1}^{N} \frac{\delta_i}{N} = \text{E}[Y(1) - Y(0)] = \text{E}[Y(1)] - \text{E}[Y(0)],$$

where E(·) is the expected value operator, or simply the mean. Note that in general neither the ICE nor the ACE are estimable as we typically only observe a unit (individual) in one of the two treatment states.

While they unpack this relationship in a much more nuanced way, the basic causal structure that Aksoy and Gambetta build on is the one between exposure to modernity, the treatment, and veiling as a form of religious expression, the outcome. The potential outcomes $Y_i(1)$ and $Y_i(0)$ indicate whether a woman wears a veil under high and low conditions of modernization. From here, δ_i stands for the difference in her behavior under these different conditions. The

---

2 Note that these quantities are sometimes referred to as Y_{ti} and Y_{ci}, or Y^t and Y^c, respectively.

3 As will be discussed, in certain cases the interest is in the effect of the treatment on the treated, the ATT, as opposed to the population as a whole. Typically, this is the case when individuals who have been selected for treatment are those for whom it is believed that the treatment will be beneficial.

King Makovi and Christopher Winship - 9781789906851

Table 20.1 Treatment assignment and outcomes observed

	Outcome	
Assignment	Treatment	Control
Treatment	$E[Y(1)\|T = 1]$	?
Control	?	$E[Y(0)\|T = 0]$

average of these individual effects over a population of women would amount to the average causal effect, i.e., how much religious expression would decrease in a society or a community, having gone through a process of modernization.

Implicit in the formula for the ACE is the strong assumption of SUTVA, the "Stable Unit Treatment Value Assumption," which states that the effect of a treatment on a unit is unaffected by how many or which other units are treated. This is obviously a strong assumption that is unlikely to hold in many situations of interest. For example, whether one veils is likely to be affected by the choices of others whom one is in contact with. Aksoy and Gambetta show that this dependence can be more complicated than simple peer influence, where the behavior of one's peers impacts one's own in the same direction.

To expose the fundamental inferential problem that the potential outcome model brings to the fore, consider the structure of data that researchers observe in terms of treatment assignment and outcomes in Table 20.1. For individuals in the treatment group, we do not know what their outcome would have been if they had been assigned to the control group. And vice versa, for individuals in the control group; we do not know what their outcome would have been if they had been assigned to the treatment group. If we did, we could calculate each individual's ICE and, from the ICEs, we could calculate the ACE by averaging the ICEs. Instead, we observe each individual under one condition but not the other.

What this means in the context of veiling is that for women living in a modern setting, we do not know whether they would veil in a setting that was less modernized and, conversely, for women living in a less modernized setting, we do not know whether they would veil in a modern setting. In short, we only know what someone would do in their actual situation and have "missing information" about what they would do in the counterfactual case.

Most estimates of an average treatment effect are some variant of what is known as "The Standard Estimator," S*, defined as

$$S^* = E[Y(1)\,|\,T = 1] - E[Y(0)\,|\,T = 0] = \overline{Y(1)} - \overline{Y(0)},$$

the difference between the average outcome for those in the treatment condition and the average outcome for those in the control condition. Note, that the two means here are each defined not with reference to the population as is true of the ACE, but with respect to two disjoint groups: the actual treatment and control groups.

There are two conditions that together are sufficient for the standard estimator (S*) to be an unbiased and consistent estimate of the ACE:

$$\overline{Y(0)} = E[Y(0)\,|\,T = 1] = E[Y(0)\,|\,T = 0] \text{ and}$$
$$\overline{Y(1)} = E[Y(1)\,|\,T = 1] = E[Y(1)\,|\,T = 0].$$

King Makovi and Christopher Winship - 9781789906851

In terms of Table 20.1, these two conditions simply amount to filling in the two question marks with the value that is observed in the same column.

Random assignment of the treatment would ensure that the two conditions defined by the two equations held in expectation. However, random assignment is often impossible, unethical or prohibitively costly. The weaker condition of ignorability is sufficient for the conditions to hold. More formally,

$$Y(1), Y(0) \quad \perp \quad T.$$

In words, potential outcomes, $Y(1)$ and $Y(0)$ (not the observed Y), must be jointly independent of treatment assignment.

In the case of Aksoy and Gambetta's paper this means that within the overall population of women under consideration the probability that a woman is in a modern context or a context that is not modern is independent of what her outcome would be in either case. This would be violated, for example, if women who wouldn't wear a veil in either context were more likely to be found in modernized contexts.

If ignorability does not hold at least approximately, then the standard estimator is likely to be a poor estimate of the ACE. We can understand the source of the difference that is the bias between the standard estimator and the true ACE in terms of the following formula:

$$S^* = \bar{\delta} + \langle E[Y(0) \mid T = 1] - E[Y(0) \mid T = 0] \rangle + (1 - \pi)(\bar{\delta}_t - \bar{\delta}_c),$$

where $\bar{\delta}$ is the ACE, the quantity we aim to estimate, $\langle E[Y(0)|T = 1] - E[Y(0)|T = 0] \rangle$ is the baseline difference in outcomes, π is the proportion of individuals in the treatment group, and $(\bar{\delta}_t - \bar{\delta}_c)$ is the difference in the effect of the treatment for the treatment and control groups, often what is referred to as selection bias (Morgan and Winship, 2015). As discussed later, a sensitivity analysis allows a researcher to assess how vulnerable their findings are to failure for exact ignorability to hold.

What is important to see here is that there are two quite separate sources of bias. First, there is the bias that most researchers consider—that individuals are initially different in some way. For example, differences in social economic background may affect the likelihood that a woman veils. Second, the effect of the treatment may vary across individuals; that is, there is treatment effect heterogeneity. This is a core part of Aksoy and Gambetta's argument: in modern societies it is more advantageous for highly religious women to veil either as a commitment strategy or to signal their piety. One of the ways that the potential outcome model has been particularly important is in making researchers aware of this second source of bias, that is, the potential importance of effect heterogeneity.

In sum, considering whether a particular estimate represents a causal effect, one first ought to ask whether all initial differences between the treatment and control groups have been controlled for. Second, one ought to consider whether it is reasonable to believe that the effect of the treatment on average is the same for the treatment and the control groups. In many cases, the answer to both of these questions will be no.[4] Pearl's DAGs will be particularly

[4] To make things worse, a seldom recognized fact is that in the presence of effect heterogeneity, OLS regression generally does not estimate the ACE (Angrist, 1998; Angrist and Pischke, 2008; Elwert and Winship, 2010).

King Makovi and Christopher Winship - 9781789906851

important to expose these shortcomings, and their analysis hold clues as to how to overcome them.

4. DIRECTED ACYCLIC GRAPHS

Although the potential outcome model specifies the formal conditions that are required for an estimate to merit a causal interpretation (ignorability), it fails to provide the conceptual tools for assessing whether any empirical situation other than a randomized experiment meets this condition. DAGs help accomplish this. First, one must posit a theoretical model for how one thinks the world works, which, of course, is open to debate—and arguably where some of the most interesting social science happens. Given a model, Pearl's DAG calculus provides a method for determining whether a particular causal effect is identified, that is, whether the only possible way for the data to be consistent with the posited theory is for the hypothesized causal effect to hold.

A particular strength of DAGs is that they are nonparametric, allowing the researcher to generalize beyond the traditional linear model (Pearl 2000, 2009). DAGs can represent arbitrarily complicated sets of equations, not just linear ones, as long as their structure is directed and contain no loops, i.e., if there is no feedback. Importantly, this provides a straightforward way of treating interactions among variables, frequent in analytical sociology.

The basic strategy in using DAGs is to explicitly represent all the relationships that might be causal and to then analyze the graph to determine which effects are identified, that is, to determine if the only way for the observed associations to be consistent with the posited theory is in fact for the hypothesized causal effect to be present.

Consider the DAG in Figure 20.1. This representation makes it clear what causal structure is assumed when considering the relationship between modernization and veiling as a form of religious expression (Aksoy and Gambetta, 2016). Specifically, by including the mediating mechanisms of signalling (S) and commitment (C), the interaction between religiosity (R) and modernization (M) becomes explicit.

Now, consider the DAG in Figure 20.2. This DAG makes it explicit that the stigma (ST) attached to certain acts influences people's likelihood of engaging in them. In the case of abortion, this means that the stigma attached to having an abortion may be a causal driver in making such a choice (S), which clearly influences disclosure (D): no abortion, no abortion secret. The stigma attached to abortion may also influence one's confidant's acceptance of an abortion secret, hence the directed arrow between (ST) and (A_A), which, in turn, might impact one's decision (S). Most importantly, however, Cowan (2014) draws our attention to how one's confidant's

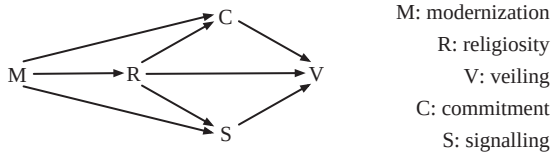


Figure 20.1 DAG representing a model of veiling proposed by Aksoy and Gambetta (2016)

King Makovi and Christopher Winship - 9781789906851

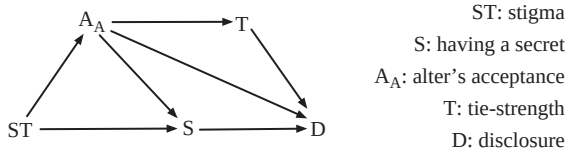


Figure 20.2 DAG representing a model of disclosure proposed by Cowan (2014)

acceptance[5] may impact disclosure directly, and how this interacts with the strength of one's relationship (T)[6] to the confidant. This is the part of the DAG that makes the mechanism of selective disclosure explicit and a matter of direct theoretical and empirical scrutiny.

As previously mentioned, the visual representation of a DAG does not imply any specific functional form for the posited causal relationships, and may include both a linear and non-linear combination of the variables. Specifically, in the case of selective disclosure, we could conceptualize these relationships in the following way:

$$D = f(S, A_A, T, T \cdot A_A),$$

indicating that both having a secret (a function of stigma and alter's acceptance) and one's confidant's acceptance has a direct effect on the probability of disclosure—intuitively, the more stigmatized a secret, the less likely it is to be shared, and the more accepting a confidant is, the more likely that it would be shared. Tie-strength may have a positive main effect on disclosure: first, it might be hard to keep secrets from close confidants; second, one may have greater motivation to disclose and maintain a shared reality with those whom one is closest. However, this may not be the whole story. Sharing a highly stigmatized secret with an unsympathetic confidant would also yield a deterrent, as the consequences under these circumstances could be particularly severe and could damage not only the relationship between the secret-holder and confidant, but could have spillover effects to other close relationships involving both of them. This is consistent with evidence that Small (2017) finds that paradoxically, one may seek some forms of support via weak, rather than strong ties.

5. IDENTIFICATION STRATEGIES

So far, the potential outcome model has made it explicit as to what is meant by a causal effect. DAGs have helped to visualize the causal relationships of interest. We now turn to three main strategies in the tool kit of empirical social scientists to help with identification: conditioning, the front-door criterion and instrumental variables. Considering these are necessary steps on our way to causal mediation analysis.

---

[5] To be more specific, the causal driver is the focal actor's *belief* about one's confidant's acceptance, (A_B), which may or may not be an accurate guess of A_A.

[6] Note that one's confidant's acceptance is not assumed to cause the strength of the relationship; rather, this arrow is meant to represent the presence of an interaction. There is variation in the literature in depicting interaction effects, some choosing to include variation in the arrows represented, a practice we do not follow here.

King Makovi and Christopher Winship - 9781789906851

5.1 Conditioning

We are first and foremost worried that the association between our dependent variable and some independent variable might be a function of their mutual dependence on other variables, i.e., a common cause. In this case, we talk about the relationship between *Y* and *T* being confounded, which means that the association between *Y* and *T* does not represent the causal effect of *T* on *Y*. Rather, *T* and *Y* are associated because of their mutual dependence on *Z*. We can illustrate this in terms of a DAG in Figure 20.3(a).

In the language of Pearl, we say that the association between *T* and *Y* fails to identify the causal effect of *T* on *Y* since that association could be due to *T*'s causal effect on *Y* or to *T*'s and *Y*'s mutual dependence on *Z*. Graphically, this is because there is a noncausal pathway from *T* to *Z* to *Y*. A noncausal path is any pathway containing arrows that do not unidirectionally go from *T* to *Y*. A causal effect between *T* and *Y* is said to be identified if the only explanation for an association (be it unconditional or conditional) between *T* and *Y* is that *T* has a causal effect on *Y*. The idea of noncausal paths is fancy but is often a helpful way to think of the problem of omitted variables.

Traditionally, we would discuss the causal graph in Figure 20.3(a) as a case where there is an omitted variable, *Z*, that would need to be controlled for in order to estimate the causal effect of *T* on *Y*. Controlling for a variable is equivalent to eliminating it from the DAG, since this is equivalent to stratifying our analysis, that is, doing a sub-analysis on the different values of the variable. In light of this identification strategy, we consider Cowan's paper again.

Cowan (2014) argues that the likelihood of disclosing a secret depends on the secret-holder's belief on how accepting their confidant will be of their secret. This is a particularly delicate issue with secrets that are stigmatized, such as having an abortion. Cowan notes that one's beliefs about how charitable a confidant may be toward an abortion secret may confound the relationship between having the secret and disclosing it. What is more, abortion secrets themselves are not randomly assigned to women, and these secrets themselves may reflect one's own acceptance of abortion in the first place. If this were the case, some woman would be unlikely to have abortions, and thus abortion secrets, and this may be particularly so in cases where women, had they had an abortion secret, would not have disclosed it. Short of data on one's own and one's confidants' beliefs, Cowan demonstrates that various demographic characteristics of those with abortion secrets compared to those without are quite similar. She argues that for correlated errors of confidant's beliefs and abortion secrets to play a confounding role, they would need to be correlated in a way that is not captured by the demographic characteristics she identifies and controls for. In essence, she argues that Figure 20.3(a) is not a good representation of her empirical case, while the DAG in Figure 20.3(b) (after controlling for various demographic factors) is.

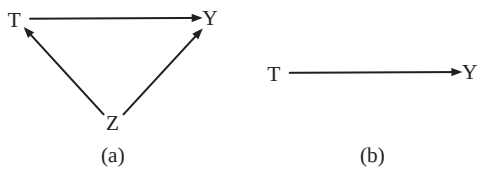


Figure 20.3 DAG illustrating a causal relationship between T and Y that is confounded by Z (a) and a DAG illustrating a causal relationship between T and Y without a confounder (b)

King Makovi and Christopher Winship - 9781789906851

If we control for *Z* in Figure 20.3(a), we get Figure 20.3(b). Now there is no noncausal path connecting *T* and *Y*. Thus, we say that the association between *Y* and *T* identifies the causal effect of *T* on *Y*. By identification, we mean that there is no other explanation for *T*'s and *Y*'s association than that of *T* causing *Y*. We now need to state our thinking more formally:

> *Theorem*: The association between *T* and *Y* consistently estimates *T*'s causal effect on *Y*, if there are no *unblocked* noncausal paths connecting *T* and *Y*.[7]

The first thing this theorem tells us is that if there are no noncausal paths connecting *T* and *Y*, we are set. It, however, goes further, saying that the only thing that we need to worry about are unblocked noncausal paths.[8] To better understand this, we need to explain what differentiates a blocked from an unblocked path.

5.1.1 Colliders

To understand what a blocked noncausal path is, we first need to understand another concept: the idea of a collider variable. A variable is a collider with respect to a particular path if there are two other variables along the path that are causes of the collider (visually speaking, two arrows pointing to the collider). In the DAG in Figure 20.4, there is a noncausal path connecting *Y* and *T*. However, *Z* is a collider variable since it is affected by both *Y* and *T* in the noncausal path. As a result, it "blocks" this pathway. As such, this noncausal path does not pose a threat to identification as it does not contribute to the association between *Y* and *T*. Putting it another way, *Z* "happens" after *T* and *Y*. As such, *Z* could in no way affect the association between *T* and *Y*.

Conditioning on a collider variable, however, can be problematic. Specifically, if we condition on a collider variable it changes the association between any prior variables that it depends on. To illustrate the problem, consider the DAG in Figure 20.5(a). Here, we assume that promiscuity (e.g., the number of sexual partners) and safe sex (e.g., using condoms) are not associated. Assume that people who are promiscuous and who have unsafe sex have a higher STI risk, and that there is no interaction between these two variables in their effect on STI risk (i.e., those who are both promiscuous and don't use protection do not get an extra bump in STI risk for combining these traits). Assume that STI risk is measured as either high or low, i.e., a dichotomous variable for the sake of simplicity. Further assume that we condition on STI risk by stratifying the data, say, only considering those with low risk. Now, take someone who

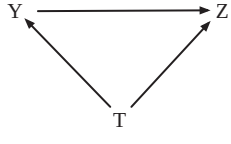


Figure 20.4 DAG illustrating a collider, Z

[7] Note that some analysts assert that a causal effect is only defined when *T* is manipulable, i.e., characteristics like gender or race do not have causal effects as long as they are thought of as non-manipulable. We assert that an alternative conceptualization of the treatment in such cases is always possible, considering comparisons where such a non-manipulable characteristic is made salient or not. An excellent example demonstrating this view is Goldin and Rouse's (2000) investigation of the impact of blind auditions on the gender composition of orchestras.

[8] To be more precise, the only thing we need to worry about is what Pearl (2000, 2009) calls a backdoor path. A backdoor path is a noncausal path with an arrow pointing at *T*.

King Makovi and Christopher Winship - 9781789906851

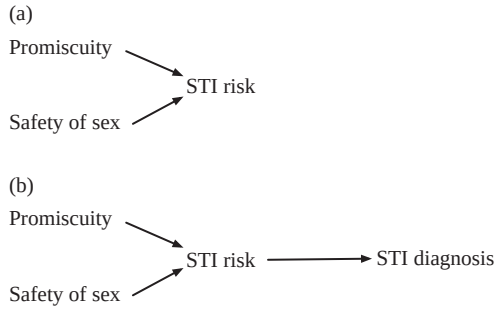


Figure 20.5 DAG illustrating the relationship between the promiscuity, the use of protection and STI risk

is promiscuous but with low STI risk. On average we would expect that they would practice safe sex. Similarly, we would expect that someone with low STI risk who is lax about using protection on average not to be promiscuous. Thus, among individuals with low STI risk, promiscuity and safe sex would be *positively* correlated. We are more likely among low STI risk individuals to have promiscuous people who use protection often, and people lax about the use of protection who are not promiscuous.

A similar relationship exists among people with high STI risk. If one is not promiscuous, then they are likely to be lax about protection. If one uses condoms often, they are likely to have sex with many partners. Again, there is a positive association. What this example highlights is that when we condition on a collider variable we need to be careful as we induce correlations between all those variables the collider was originally dependent on, even if they were originally independent.

Conditioning on the descendent of a collider results in the same problem. Consider the example in Figure 20.5(b). Descendants of a variable are variables that are caused directly or indirectly by them. STI diagnosis is a descendant of STI risk, those at high risk of having an STI are also those at high risk of being diagnosed with one. STI diagnosis, therefore, is also a descendant of promiscuity and practicing safe sex. When we condition on the descendant of a variable, it is like we partially condition on that variable, thus we end up with the same effect.

In the case of STI risk and diagnosis the relationship is clear. While not all STIs are diagnosed, as some remain without symptoms, and not everyone turns to a doctor to treat them, most STIs yield symptoms that prompt a doctor's visit and a diagnosis. A more subtle case would be to condition on fertility. Given that infertility may be caused by STIs, especially if they remain untreated, the same effect would obtain if one conditioned on it. Note that conditioning may enter in subtle ways other than controlling, such as via one's sampling strategy. Imagine, for instance a study of PrEP use (which could be an antecedent of promiscuity and condom-less sex) that recruits participants at an infertility clinic.

More generally, if we condition on a collider variable, we say that we have unblocked a path. Thus, conditioning on colliders and noncolliders have different effects:

1. A noncollider variable does not block a path, but conditioning on that variable does block the path.
2. A collider variable blocks a path, but conditioning on that variable (or its descendants) unblocks the path.

King Makovi and Christopher Winship - 9781789906851

The good news is that conditioning on a noncollider variable blocks the contribution of a noncausal path to an observed association. What is more, this blocking variable can exist anywhere along the path of concern. Unfortunately, however, collider variables are common and often undermine efforts to identify a relationship. Elwert and Winship (2014) provide an extensive discussion of substantive examples where colliders create problems, but see also recent work by Breen (2018). The basic lesson demonstrated here is that when one wants to condition on a collider a red flag should go up.

5.2 The Front-Door Criterion

Conditioning on a noncollider variable provides one strategy for achieving identification. Pearl's *front-door criterion* provides another. In addition, it formally demonstrates how a mechanism-based analysis can warrant a claim of a causal effect.

The front-door criterion identifies the causal effect of a variable on an outcome by augmenting the causal model to include the intermediate variable(s) through which that variable affects the outcome in question. If it is possible to identify the effect of the variable of interest on each of the intermediate variables, and to identify the effect of each of the intermediate variables on the outcome, then the (total) effect of the variable of interest on the outcome can be estimated as the product of the effects of the paths connecting them.

Pearl (2000) uses the classic example of the effect of smoking on cancer. In a similar vein, in Figure 20.6 we would like to estimate the effect of taking PrEP (T) on STI risk (Y). We believe that taking PrEP affects STI risk through promiscuity (M). The association between T and M does not provide a consistent estimate because of the unblocked noncausal path through U, where U represents possible behavioral (e.g., risk preferences), or environmental (e.g., the promotion of health consciousness in one's community) factors. If U is observed, then we can estimate the effect of T on Y by conditioning on U. If U is unobserved, which we represent by enclosing it in circle, this strategy cannot be pursued. However, if we can consistently estimate the effect of T on M, i.e., the effect of taking PrEP on promiscuity, and the effect of M on Y, i.e., the effect of promiscuity on STI risk, getting estimates of b and c, then we can estimate the effect of T on Y by combining the two. In the case of a linear model, the effect would be $b{\cdot}c$. This is the core idea behind the front-door criterion.

In the present case, we can estimate both b and c (or their functional equivalents in a nonlinear model) by a double application of our standard conditioning criteria (Pearl, 2000, 2009). Because there are no unblocked noncausal paths between T and M, we can consistently estimate the effect of T on M from their association. While there is a noncausal path between M and Y through T and U, conditioning on T, we can eliminate this backdoor path, which allows us to consistently estimate the effect of M on Y. As we return to this example later on,

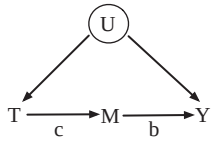


*Figure 20.6 DAG representing the effect of taking PrEP (*T*) on STI risk (*Y*) via promiscuity (*M*)*

King Makovi and Christopher Winship - 9781789906851

this may not be the only way in which PrEP status impacts STI risk, but for now this strategy provides an estimate for the impact of taking PrEP via promiscuity, a mediating mechanism, on STI risk.

The front-door identification criterion is important as it is a way of dealing with unblocked noncausal paths that cannot be eliminated because key variables that we want to condition on (control for) are not observed, thus making regression, stratification and matching on those variables impossible.

5.3 Instrumental Variables

The use of instrumental variables (IVs), a strategy often pursued by economists, is a third approach to identification alongside conditioning and the front-door criterion. The strategy here is to find a variable that only affects the outcome through the treatment[9] and is not associated with the errors or other variables confounding the relationship between *T* and *Y*, also called the exclusion or exogeneity criterion. In a way, it is not entirely dissimilar from the broad idea of the front-door criterion, as in both cases one looks to elaborate on the causal chain connecting *T* and *Y*, augmenting it "forward" or "backward."

Consider the DAG in Figure 20.7. What we see here is that *Z* meets the criterion. *Z* only affects *Y* via *T* and is not associated with the two error terms U_t and U_y (note that the double-headed arrow connecting them represents common causes of both, not explicitly visualized in the figure). The association between *Z* and *T* identifies the causal effect of *T* on *Z*, *a*, as there is a noncausal path from *Z* to *Y* going from *Z* to *T* to U_t to U_y to *Y*. However, *T*, is a collider along this path, and thus it is blocked. As such, the causal effect of *Z* on *Y* is also identified. In a linear model this would be equal to $a \cdot b$. Since we can also estimate the effect of *Z* on *T*, in the linear case $\hat{a}$, we can estimate the effect of *T* on *Y* as $\hat{b} = \frac{\widehat{ab}}{\hat{a}}$.

The causal effect of PrEP status on STI risk (excluding HIV) has long been debated. More recently, data started to be available at a high geographic resolution on access to PrEP and PrEP use, via collating administrative data on providers and prescriptions filled. Correlational analyses show that places with high levels of PrEP availability have seen increases in bacterial STIs. However, many unobservables confound this relationship, such as the size of the MSM community, the largest group of PrEP users in Western contexts. A significant improvement over these correlational investigations could be to use an instrument. There are two obvious candidates. First, the closure of existing practices, or the opening of new ones distributing PrEP for reasons unrelated to PrEP uptake in the respective community they operate in. Should such administrative changes be identified, one could

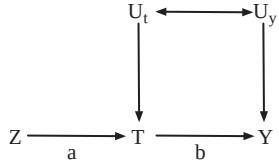


Figure 20.7 DAG representing an instrumental variables strategy for identification

[9] This condition is often referred to as "relevance" (Staiger and Stock, 1997).

King Makovi and Christopher Winship - 9781789906851

compare the growth in STI infections in places that experienced such closures/openings, influencing the availability of PrEP, to places that saw no change in PrEP availability. Second, PrEP prescription guidelines have seen multiple updates over the years since PrEP has been released, seeing a gradual expansion. Should the timing of these interventions be unrelated to previous PrEP uptake, this change could also be used to identify the causal effect of PrEP on STI diagnosis. Note, however, that in both cases the interpretation of this effect is still open to dispute as it could map onto increased testing, as well as behavioral mechanisms such as risk compensation.

A more common way to think of instrumental variables, particularly in the epidemiological and medical literature, builds on the realization that even randomized control trials are non-random. Individuals assigned to a treatment group may not stay with their prescribed regimens. Who follows through with their treatment is therefore not random; treatment assignment is. Formalizing this, Z in this case would be a so-called intention to treat, T would be treatment take-up and Y the outcome under consideration. With all these pieces of information one could follow the calculus used for estimation to back out the true treatment effect.

In summary we have three strategies for identifying a causal effect: (1) conditioning on a noncollider variable, which unconfounds the relationship between T and Y; (2) the front-door criterion; and (3) using instrumental variables. All three methods are potentially useful in mediation analysis.

6. MEDIATION ANALYSIS

Mediation analysis is concerned with assessing what portion a variable's effect on an outcome is explained by one or more mediating variables. Consider the case where there are only three variables: T, often referred to as the treatment; Y, the outcome of interest; and M, the variable that mediates some portion of the effect of T on Y, substantively, a mechanism. Note that we use "mechanism" here in the sense that it unpacks the "black box" of a treatment to outcome relationship by elaborating on "how" the causal effect is brought about (via M). This interpretation has been criticized by Manzo (2014), among others, arguing that no statistical analysis may live up to the richness of the mechanism concept as captured by analytical sociologists via agent-based models that explicitly theorize and model behaviors and interactions. While we appreciate the criticism, we assert that the conceptualization of a mechanism here is equally explicit and systematic (Hedström and Yilikoski, 2014).

The importance of the recent mediation literature is two-fold: first it shows that there is a key assumption behind traditional mediation analysis that is often unlikely to be met; second, it provides a general, possibly a fully nonparametric, approach to calculating the indirect effect in a mediation model.

Consider the DAG in Figure 20.8. Here, we see that for individual i, the effect of T on M is α_i, the effect of T on Y is β_{1i} and the effect of M on Y is β_{2i}. In the language of the traditional mediation literature, for individual i, T's direct effect on Y is DE = β_{1i}, its indirect effect on Y is IDE = $\alpha_i \cdot \beta_{2i}$ and its total effect on Y is the sum of the two: TE = $\beta_{1i} + \alpha_i \cdot \beta_{2i}$ (Baron and Kenny, 1986). An important attribute of the model in Figure 20.8 based on the traditional approach to mediation analysis is that it assumes that the effect of the treatment, T, and the mediator, M, do not interact. If they do, then the simple decomposition formula no longer applies. We discuss this situation further.

King Makovi and Christopher Winship - 9781789906851

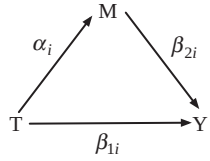


Figure 20.8 DAG representing a simple mechanism, noting individual effects

Typically, the goal of a researcher is to estimate these parameters for a population, not an individual as previously discussed in case of the ACE.[10] As we move from a single individual to a population, two potential problems arise. First, there may be heterogeneity in effects, specifically in α, β_1 and β_2.[11] Second, heterogeneous effects may be correlated with one another. This introduces a different kind of problem, as we will discuss.

The first issue may be dealt with by conditioning and is akin to a model misspecification type of problem, where estimates are no longer unbiased and consistent (e.g., Morgan and Winship, 2015). In more substantive terms, they would not identify the causal effect and thus should not be interpreted as such. If these differences are fully a function of observed variables, that is, there are observed moderators, they can (and should) be modelled using interaction terms. Of course, it is always possible that there is unobserved effect heterogeneity, with the result that some parameters are likely under-identified. In this case, the researcher is back to the drawing board and has to attempt to use some of the techniques mentioned earlier to identify the effect(s) in question, or revert to sensitivity analyses.

The second problem with heterogeneous effects is that they may be correlated with each other, specifically in the case of α and β_2. What new mediation literature has shown is that the traditional product formula for the IDE = $\alpha \cdot \beta$ does not hold in this case[12] (Imai et al., 2010a, 2011; Pearl, 2014; VanderWeele, 2015; Huber, 2019). This issue can be understood by considering the general formula for the expected value of the product of two parameter estimates, which will only equal the product of expected values if they are independent ($E[\alpha\beta_2] = E[\alpha] \cdot E[\beta_2]$ only if ($\alpha \perp \beta_2$)).

To expose this issue more fully, consider the case where we have two individuals, i and i^*, with the following parameter values reflected in Table 20.2. Here, the true IDE = -1. This, however, is not the answer we get when calculating the IDE as the product of $E[\alpha] \cdot E[\beta_2] = 0 \cdot 0 = 0$, giving us a TE = 1 instead of 0.[13] The reason behind this is that the effects α and β_2 for i and i^* are perfectly negatively correlated, and the formula here fails to take this into account.

Recall the example of PrEP status on STI risk and add a slightly more complex causal structure, where the impact of PrEP status is also mediated by STI testing in Figure 20.9. This is a pathway that received a lot of coverage in recent literature. Without tests, the disease cannot be identified, and this way testing figures prominently in our understanding of disease spread.

---

[10] Since we generally only observe an individual in one of the two treatment statuses, it is often not possible to estimate individual effects, and often times not even desirable, which is not to say that the same would not be particularly informative for groups of individuals.

[11] Of course, if effects are constant across individuals, the α, β_1 and β_2 apply to everyone, the subscript i is no longer needed and the problems described do not arise.

[12] Baron and Kenny (1986) did indicate that the traditional IDE formula did not hold in the linear case if errors across the two equations were correlated. This point has almost been universally ignored.

[13] Imai and colleagues (2011) provide a more complicated version of this example.

King Makovi and Christopher Winship - 9781789906851

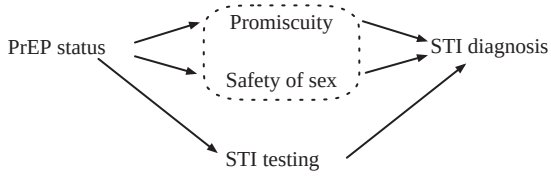


Figure 20.9 An example of correlated heterogeneous effects. Note that the dotted circle highlights the mechanisms of risk compensation and differentiates them from STI testing

Table 20.2 An example of correlated heterogeneous effects

	α	β_1	β_2	$\alpha \cdot \beta_2$	$\beta_1 + \alpha \cdot \beta_2$
i	1	1	–1	–1	0
i^*	–1	1	1	–1	0
$E[\cdot]$	0	1	0	0	1

Current medical guidelines mandate regular HIV testing for those who take PrEP, and at these occasions free STI testing is often offered. Arguably, this way some cases get diagnosed that otherwise would remain undetected, for instance, the ones that are asymptomatic.

Now, imagine two groups of people. The first gets tested more often as a result of taking PrEP than they would have otherwise, which, somewhat counterintuitively, might lower their chances of an STI diagnosis. Some argue that testing in and of itself might make people more health and hygiene conscious, and would foster the development of better habits of avoiding STIs in general. Furthermore, on the occasions of getting tested, materials on STI spread and prevention are often made available, further contributing to lowering the chances of STI infections. For this group, the impact of PrEP via STI testing will be negative, as they may develop better health habits and contract STIs less frequently as a result of heightened testing and heightened health consciousness.

The second group gets tested less often, as their primary reason for frequent testing previously may have been their concern of contracting HIV, which they virtually are no longer at risk of (provided that they take PrEP as prescribed), while, when they do get tested, it is a result of STI symptoms other than HIV driving STI diagnosis. For this group, the impact of PrEP via STI testing will be negative as well due to the fact that asymptomatic cases would get caught less often than it would have been the case without introducing PrEP.

If the effects described are equally strong, and these groups are of similar sizes, the case described here is identical to the one represented in Table 20.2, and showcases how effect heterogeneity and correlated effects may lead to erroneous conclusions when one ignores them.

6.1 A General Approach to Mediation Analysis

In light of the complexities outlined so far, an approach to mediation analysis is needed that allows for any type of relationship between variables, avoiding assumptions of linearity and

King Makovi and Christopher Winship - 9781789906851

the absence of interactions. Here, we show that the potential outcome framework accomplishes this.[14]

In order to simplify the exposition further, without loss of generality, assume that all variables are dichotomous. Consider the DAG in Figure 20.10 and the overarching goal to estimate the effect of the treatment (T) via the mediating mechanism (M) on an outcome of interest (Y). Note that there are the two potential outcomes for M: $M_i(1)$, and $M_i(0)$ and four potential outcomes for Y: $Y_i(1, 1)$, $Y_i(1, 0)$, $Y_i(0, 1)$, $Y_i(0, 0)$, representing the four possible combinations of values for T and M. Using the potential outcome framework, we can define the effect of T on M, the effect of T on Y and the effect of M on Y for individual i as

$$
\begin{aligned}
&T_i\text{'s effect on } M_i,\ \alpha \quad : M_i(1) - M_i(0),\\
&T_i\text{'s effect on } Y_i,\ \beta_{1m} \ : Y_i(1,m) - Y_i(0,m) \text{ for } m = 0,1,\\
&M_i\text{'s effect on } Y_i,\ \beta_{2t} : Y_i(t,1) - Y_i(t,0) \text{ for } t = 0,1.
\end{aligned}
$$

Note that T's effect on M does not vary with the value of other variables. It is equivalent to α in the DAG in Figure 20.10, though remember α may differ across individuals. The effect of T on Y (which can also be individual-specific) potentially depends on M and, similarly, the effect of M on Y (again, which can vary across individuals) potentially depends on T. Thus, this formulation allows T and M to interact. Note also that there is no assumption of a functional form for combining the effects of T and M, such as linearity. Consequently, this formulation is general and nonparametric.

Substantively, now consider the impact of PrEP (T) on STI diagnosis other than HIV (Y), and the mediator, promiscuity, measured by the number of sexual relationships (M) over a given period of time. T_i's effect on M_i is the difference in the number of partners one may have on PrEP compared to off PrEP. M_i's effect on Y_i is the effect that the number of partners have on STI diagnosis, interpreted both on and off PrEP. Note that these two quantities may or may not be the same. It is possible that people choose different partners on and off PrEP, partners who potentially carry different STI risks. Specifically, MSM on PrEP may choose more experienced partners who have had more partners themselves, while those not on PrEP look for partners with few prior partners. This example exposes the possible interaction of the treatment and mediator explicitly. Finally, T_i's effect on Y_i is the impact of PrEP on STI diagnosis.

In addition to the three core potential outcome equations, we can use the potential outcome framework to nonparametrically define the three quantities that are bread and butter to formal mediation analysis: the direct effect of T on Y (DE), the indirect effect of T on Y (IDE[15]) and the total effect of T on Y.

$$
\begin{aligned}
&\text{Direct Effect (DE)} \quad : Y_i(1, M(t)) - Y_i(0, M(t)) \text{ for } t = 0,1,\\
&\text{Indirect Effect (IDE)} : Y_i(t, M(1)) - Y_i(t, M(0)) \text{ for } t = 0,1,\\
&\text{Total Effect (TE)} \quad : Y_i(1, M(1)) - Y_i(0, M(0)).
\end{aligned}
$$

[14] Rubin has been critical of formal mediation analysis as presented here, albeit it builds on the framework he himself made major contributions to. He evidently recognized the role that mediators, or post-treatment covariates, play in bringing about outcomes and proposed a principle stratification framework as a way to define causal effects in this case (Frangakis and Rubin, 2002).

[15] Not to be confused with interventional direct effect as in Nguyen et al. (2020). It is also referred to as IE.

King Makovi and Christopher Winship - 9781789906851

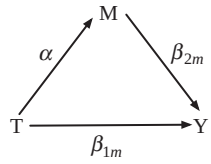


Figure 20.10 DAG representing a simple mechanism, noting homogeneous effects

Note in the equation for the direct effect, whatever value t is set to determines the value that M takes: $M(t)$, i.e., an individual's $M(t)$ is induced by their value on t. This is what is known as the natural direct effect. One could imagine instead separately setting M to some value m and then evaluate the effect of changing T on Y. This is what is known as the "Controlled Direct Effect" (CDE). In this chapter, we will only be concerned with "Natural Direct Effect," and will simply refer to it as the "Direct Effect" (DE).

As strongly emphasized by Pearl (2000, 2009) throughout his work, and Bollen and Pearl (2013), the outcome functions are structural, not statistical. For example, in the case of Y, $Y_i(t, m)$ represents the value that Y would take on setting $T = t$ and $M = m$. This is to say, as of yet, nothing has been said about how to use data to estimate these effects.

6.1.1 Assumptions for identifying mediation effects

One of the key contributions of the recent literature on mediation analysis is to make clear what conditions are needed to identify mediation effects, particularly the IDE. Consider the DAG in Figure 20.10, which now is for a population of individuals as opposed to a single person. Our initial goal is to identify the five parameters in this DAG. Note that there are two β_1s since β_1 varies with m and there are two β_2s since β_2 varies with t. In addition, we would like to identify the DE, IDE and TE. Following Pearl (2014), four assumptions are needed:[16]

1. If the causal relationship between T and M is confounded by other variables, it must be possible to identify the relationship by some means: conditioning on a variable(s), the front-door criterion, or using instrumental variables.
2. If the causal relationship between T and Y is confounded by other variables, it must be possible to identify the relationship by some means: conditioning on a variable(s), using the front-door criterion or using instrumental variables.
3. If the causal relationship between M and Y is confounded by other variables, it must be possible to identify the relationship by conditioning.
4. No variable that is conditioned on can be affected by T.

Assumptions (1) and (2) are conventional. Simply put, if a causal relationship is confounded one needs to use one of the identification strategies discussed earlier in this chapter. Assumption (3) is closely related to assumptions (1) and (2). If M and Y are confounded, they need to be deconfounded. But why can't either the front-door criterion or instrumental variables be used? This relates intimately to why the standard formula for the indirect effect,

[16] In a series of papers, Imai and colleagues (2010a and b, 2011) consider a more restrictive set of conditions for identification called sequential ignorability. We follow Pearl (2014), both because his assumptions are more general and because they are easier to understand and apply, while we implement slight modifications and a different order.

King Makovi and Christopher Winship - 9781789906851

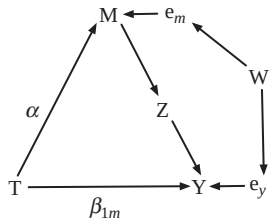


Figure 20.11 DAG representing a mechanism, and an unobserved, omitted variable (W) that impacts the mediator and the outcome

$\alpha \cdot \beta_2$, may provide a biased, even highly biased, estimate of the actual indirect effect when α and β_2 are correlated. As discussed, calculating the indirect effect for an individual involves multiplying α_i with β_{2i}. In order to get the IDE for the population, we need first to calculate the product for each individual and then take the mean of these products. The one case where this does not cause a problem is when α and β_2 are independent, which may or may not apply to an empirical case.

Consider the DAG in Figure 20.11. Think of e_m as representing individual variation in α_i and e_y as representing individual variation in β_{2i}.[17] Let W be a confounding variable that creates an association between the two errors, and thus an association between α and β_2, and more generally confounding the relationship between M and Y. It would seem like the variable Z would allow one to use the front-door criterion to get around the fact that M and Y are confounded, allowing us to consistently estimate β_2. Recall however, that this strategy would hinge on our ability to estimate both the effect of M on Z, and the effect of Z on Y, the latter of which would be used for conditioning on M to block an unblocked backdoor path, which could be accomplished successfully, save for the fact that M now is a collider on a different path, opening up a path of association that has been previously closed. This could only be resolved by conditioning on W, which is unobserved. As such, the two errors are still not independent as needed to calculate the IDE.

Unfortunately, there is an even more subtle issue with assumption (3). More formally, we could write this assumption as stating that $Y(t, M(t)) \perp M(t')$, where t and t' are not necessarily equal. This assumption is needed due to the fact that both DE and IDE contain terms where $t \neq t'$. For example, in the case where $t = 0$, the first term in the DE is $Y_i(1, M(0))$. Note that this implies that for person i, $T_i = 1$ and 0 at the same time, generally an impossibility. Under such circumstances, there are no observed individuals with which to *directly estimate* this quantity. The independence assumption allows us to use observed individuals whose potential outcomes are realized values of $Y_i(1, M(1))$ or $Y_i(0, M(0))$ to estimate the $Y_i(1, M(0))$ term in the DE. For more details, see Imai et al. (2010a and b, 2011); Pearl (2014); Andrews and Didelez (2020).

Assumption (4) is about avoiding what are known as "damned if you don't, dammed if you do" situations. Consider the DAG in Figure 20.12.[18] If we don't condition on N, then N confounds the relationship between M and Y. Note, however, that T has a direct effect on M and an indirect effect through N. If we condition on N in order to eliminate the confounding,

[17] To be precise, M's error term for an individual i is $e_m + T \cdot \alpha_i$. Similarly, for e_y.
[18] Adapted from Imai and colleagues (2011).

King Makovi and Christopher Winship - 9781789906851

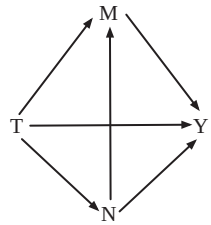


Figure 20.12 DAG representing a mechanism, and a more complex causal structure

we will block part of the effect of *T* on *M* and thus part of *T*'s indirect effect on *Y* that goes through *M*.

More formally, we have that the IDE is defined as

$$\text{IDE}: Y_i(t, M(1)) - Y_i(t, M(0)) \text{ for } t = 0,1.$$

Note the IDE is evaluated at that level of *M* that is induced by *T*. If we control for *N*, we would eliminate some of the effect of *T* on *M* and thus some of *T*'s indirect effect on *Y*. In general, variables that are descendants of *T* and also causes of *M* create damned if you do, damned if you don't situations. Assumption (4) requires that there be no such variables.

Assumption (4) is only required to assure nonparametric identification. A current area of research is the development of methods to deal with damned if you do, damned if you don't situations for parametric models. See Morgan and Winship (2015, pp. 392–415), references there within, and recent work by Wodtke and Zhou (2020) and Zhou and Wodtke (2020). A discussion of these methods is beyond the scope of this chapter.

6.2 Estimation of the Causal Mediation Effect (IDE)

A large literature has considered the problem of estimating the IDE in nonlinear models and when variables interact. Imai and colleagues (2011) show how the potential outcome model provides a general and straightforward way for calculating the IDE whether the model is nonlinear and/or contains interactions.

Previously we discussed various identification strategies that can be used as a foundation for estimating the basic causal parameters in a mediation model. A researcher may use linear regression with interactions, logit, probit or some other type of general model, or take a fully nonparameteric approach (provided that there is enough data). What Imai and colleagues (2011) show is that one can simply use the resulting equations to predict the full set of potential outcomes for each individual, then calculate each individual IDE, which then enables calculation of the IDE for the whole population (for a more technical treatment, see Huber (2019)). Note, though, that it must be the case that α_i and β_{2i} are conditionally independent or equivalently that *M* and *Y* are not confounded since, as discussed, there is no reason to believe that otherwise the estimates are unbiased and consistent. Statistical software in R has been developed by Imai and colleagues (2010c) to perform estimation.

Unfortunately, the critical assumption of no confounding of *M* and *Y* is not empirically testable. It must be made and argued for on theoretical grounds. Imai and colleagues (2010a,b) develop a method for analyzing how sensitive results are to this assumption. They have

King Makovi and Christopher Winship - 9781789906851

two implementations. In the first, the user specifies the correlation between e_m and e_y (see the DAG in Figure 20.11). This determines the bias in the estimate of the IDE. In the second, in order to enhance interpretablity, the user specifies how much the R^2 will change if one could control for the omitted variable *W* in the same DAG.

6.3 Extended Empirical Example

The past couple of years have seen a heated debate on the behavioral impact of taking PrEP on sexual health outcomes via the mechanism of risk compensation (Calabrese et al., 2017; Harawaa et al., 2017; Coyer et al., 2018; Newcomb et al., 2018; Chen et al., 2019), which is a cognitive behavioral process when individuals take more risk based on their belief that they are protected from adverse consequences (Eaton and Kalichman, 2007; Hogben and Liddon, 2008). The literature identifies two mechanisms of risk compensation as a result of taking PrEP: increased promiscuity and decreased preferences for safety. An often mentioned major threat to identification is increased testing for STIs for those on PrEP,[19] which is not a mechanism of risk compensation, but also mediates the effect of taking PrEP on STI diagnosis. Some previous work criticized others for interpreting the effect of PrEP status on STI diagnosis without controls for STI testing frequency, and rightfully so. Here, we demonstrate that using the front-door criterion, and explicitly theorizing risk compensation as a mediating mechanism, helps with identification, and sensitivity analyses allows us to gauge our confidence in the results.

In order to identify the impact of risk compensation on STI diagnosis, a number of characteristics need to be controlled for that induce an association between taking PrEP and sexual health outcomes that are non-causal. These fall in two broad categories: demographic characteristics, e.g., socio-economic status, and behavioral factors such as risk preferences. Figure 20.13 represents these as a DAG, expanding on parts of this example that have already

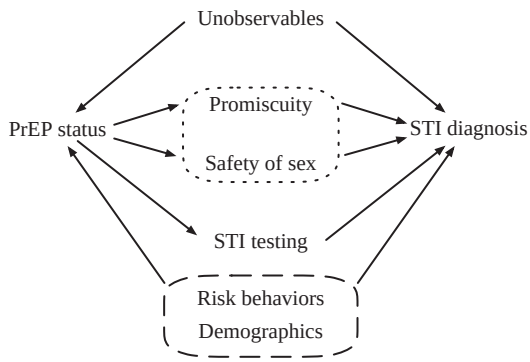


Figure 20.13 DAG representing the causal effect of taking PrEP via two mechanisms associated with risk compensation—promiscuity and safety of sexual contact—on STI diagnosis. Note that the dotted circle highlights the mechanisms of risk compensation and differentiates them from STI testing; the dashed line highlights observed controls

[19] While on PrEP, frequent STI testing is recommended as per the CDC guidelines: https://www.cdc.gov/hiv/pdf/risk/prep/cdc-hiv-prep-guidelines-2017.pdf.

King Makovi and Christopher Winship - 9781789906851

been introduced. As previously articulated (see Figure 20.11), the key to identification is the lack of presence of unobservable variables impacting the mechanism(s) of risk compensation and the outcome. Recall that unobservable factors that impact one's PrEP status and STI diagnosis are no threat to identification when it comes to mechanisms of risk compensation (see Figure 20.6).

The analyses presented here are based on data collected between January and March of 2020 in New York, London, Sydney and Toronto.[20] Participants were reached via Facebook's advertising services that led to the survey for the study. Those who reported that they were HIV-negative men who have sex with men living in these metro areas, and were above 18, were allowed to participate in the study and were included in the analyses reported here if they had completed the relevant part of the questionnaire: 1912 individuals.

All measures described, including STI diagnoses, are based on self-reports. PrEP status is measured as having ever taken PrEP, promiscuity is measured by two variables, the number of sexual partners over the past 12 months and being in multiple ongoing relationships (i.e., concurrency), and safe sex is measured with the share of times one has used condoms over the past 12 months (ranging from 0% to 100%). STI diagnosis includes any diagnosis over the past 12 months of gonorrhoea, chlamydia or syphilis—the most common bacterial STIs primarily impacting MSM. Testing frequency is the number of times one got tested over the past two years. Importantly, we measure risk behaviors with multiple measures as well: we use a behavioral measure of risk aversion (Eckel and Grossman, 2002), a measure of using hook-up sites for finding casual sex partners, and substance use over the past 12 months. Demographic control measures include metro area, age, race (majority/minority status for purposes of consistency), employment status, coarsened income percentile ranking and education. From the analytical sample, 72 respondents were dropped who chose to withhold information on their PrEP status, drug use or their use of hook-up sites or apps, yielding 1840 respondents.

We perform causal mediation analysis using the `mediation` package in `R` (Imai et al., 2010c; Tingley et al., 2014). Estimation and sensitivity analyses are based on general algorithms developed by Imai et al. (2010a, b). As previously described, estimation works by specifying a mediator model (for the effect of the exposure on the mediator), and an outcome model (for the effect of the mediator on the outcome, while controlling for the exposure), and calculating the effect of the exposure on the outcome combining the two relevant estimates. In this present case, the exposure is PrEP status, the mediators of interest are the mechanisms of risk-compensation, and the outcome is any STI diagnosis received. For simplicity, we present these as linear structural equation models as in equations (20.1) and (20.2) in the updated version of (Tingley et al., 2014) without the interaction effects between the mediator and treatment:[21]

$$M_i = \alpha_2 + \beta_2 T_i + \xi_2^T \mathbf{X}_i + \varepsilon_{i2}, \tag{20.1}$$

$$Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \xi_3^T \mathbf{X}_i + \varepsilon_{i3}. \tag{20.2}$$

[20] The original research was conducted by Kinga Makovi and Filip Karan. Data, analyses and results are discussed more extensively in *The Impact of Taking Pre-exposure Prophylaxis on Risk Compensation of Men Who Have Sex with Men*, available upon request. The data collection presented here was approved by the NYU Abu Dhabi IRB.

[21] https://cran.ism.ac.jp/web/packages/mediation/vignettes/mediation-old.pdf

King Makovi and Christopher Winship - 9781789906851

Table 20.3 Results of the casual mediation analysis following the DAG presented in Figure 20.13. The confidence intervals are based on nonparametric bootstrap with 1000 resamples

	Estimate	95% CI	p-value
Number of partners			
TE	0.171	[0.131, 0.210]	< 0.001
IDE	0.005	[0.002, 0.020]	0.010
DE	0.166	[0.124, 0.200]	< 0.001
% Mediated	0.030	[0.009, 0.100]	0.01
Concurrency			
TE	0.171	[0.132, 0.210]	< 0.001
IDE	0.008	[0.002, 0.020]	0.002
DE	0.163	[0.124, 0.200]	< 0.001
% Mediated	0.048	[0.015, 0.090]	0.002
Condom use			
TE	0.172	[0.130, 0.210]	< 0.001
IDE	0.006	[< 0.000, 0.010]	0.040
DE	0.166	[0.124, 0.200]	< 0.001
% Mediated	0.036	[0.002, 0.080]	0.04
STI testing			
TE	0.172	[0.132, 0.210]	< 0.001
IDE	0.050	[0.036, 0.070]	< 0.001
DE	0.122	[0.081, 0.160]	< 0.001
% Mediated	0.292	[0.209, 0.400]	< 0.001

Note, however, that this formulation is not appropriate, as the outcome—whether or not someone received an STI diagnosis—is a binary variable, which we model as a probit regression. Note additionally that for three out of the four mediator models, a linear regression is appropriate (number of partners,[22] likelihood of condom use and frequency of STI testing), while for concurrency—a binary variable—a probit model is more appropriate.[23]

The results of these analyses are in Table 20.3. One can see that the proportion of the effect that is mediated by the behavioral mechanisms consistent with risk compensation amount to

[22] Currently, negative binomial regression is not available, which would be a more appropriate functional form, as the number of partners reported is a highly skewed distribution. Robustness analyses show, however, that dropping the few individuals who report many more partners than the rest of the sample does not alter the results.

[23] Currently, the function `mediations(.)`, which provides simultaneous estimation for multiple mediation models allows one to select a functional form that applies to all mediator models at once. For this reason, we estimate separate models for each mediator, which is what we report. The results are virtually identical when considering concurrency as a linear probability model, allowing for the same specification for all mediator models.

King Makovi and Christopher Winship - 9781789906851

Table 20.4 Results of the casual mediation analysis reported on the treated only following the DAG presented in Figure 20.13. The confidence intervals are based on nonparametric bootstrap with 1000 resamples

	Estimate	95% CI	p-value
Number of partners			
TE	0.171	[0.131, 0.210]	< 0.001
IDE	0.007	[0.002, 0.020]	0.010
DE	0.168	[0.125, 0.200]	< 0.001
% Mediated	0.041	[0.012, 0.130]	0.010
Concurrency			
TE	0.171	[0.132, 0.210]	< 0.001
IDE	0.011	[0.003, 0.020]	0.002
DE	0.166	[0.126, 0.200]	< 0.001
% Mediated	0.065	[0.021, 0.120]	0.002
Condom use			
TE	0.172	[0.130, 0.210]	< 0.001
IDE	0.008	[< 0.000, 0.020]	0.040
DE	0.168	[0.126, 0.200]	< 0.001
% Mediated	0.049	[0.003, 0.100]	0.040
STI testing			
TE	0.172	[0.132, 0.210]	< 0.001
IDE	0.065	[0.048, 0.080]	< 0.001
DE	0.136	[0.091, 0.170]	< 0.001
% Mediated	0.378	[0.280, 0.490]	< 0.001

about 11.5%[24] of the total effect, while the portion associated with STI testing is roughly three times as large. These analyses reveal that 60.0% of the effect of being on PrEP on STI diagnoses is not accounted for by any of these mechanisms. Table 20.4 reports the same on the treated, i.e., on those who are on PrEP. Note that for this group risk compensation accounts for 15.5% of the effect, while STI testing for about 38.0%, i.e., only over 40.0% of the total effect remains theoretically unexplored.

Sensitivity analyses allow us to investigate what the approximate range of ρ is where the sign of the causal mediation effect is indeterminate. When considering the number of partners, this is at the value of 0.06,[25] i.e., even low levels of correlation between the errors in the mediator and the error in the outcome would yield a mediation effect that is zero. Alternatively, sensitivity analyses also supply the value of $R_M^{*2} \cdot R_Y^{*2}$, which represents the proportion of the

[24] This rough estimate corresponds to theorizing no interaction effects between mediators. The aggregate should not be interpreted as a causal effect, while individual effects, provided that identifying assumptions hold, can be.

[25] The sensitivity analysis function reports values separately for the treatment and control groups. The ones referenced in the text are the values for the treatment group, while they are almost identical to those in the control group.

King Makovi and Christopher Winship - 9781789906851

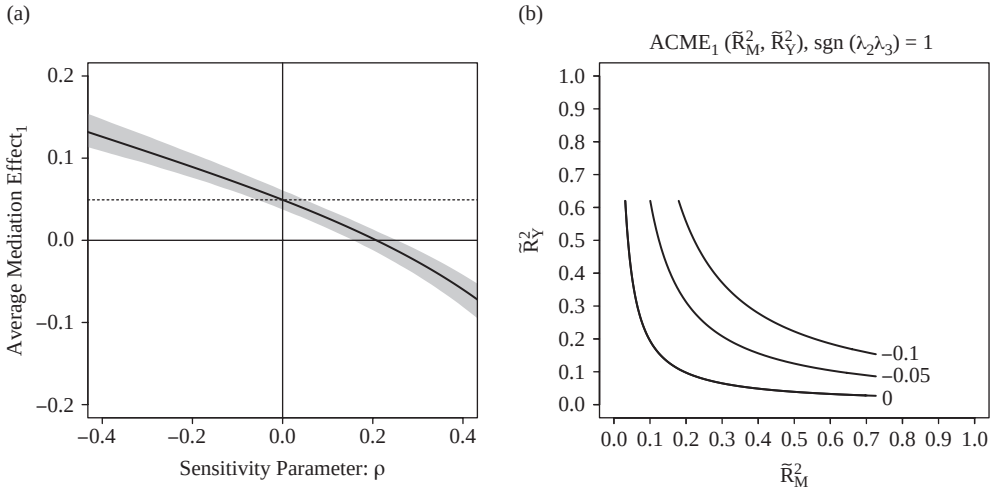


Figure 20.14 Sensitivity analysis for the mediation effect of testing frequency

previously unexplained variance in the mediator and outcome variables that is explained by an unobservable pretreatment confounder. In this case the corresponding value is 0.004. Another way to think of this quantity is in terms of the proportion of the *original variance explained* by an unobserved confounder, denoted by $\widetilde{R}^2_M \cdot \widetilde{R}^2_Y$, amounting to 0.002 (Imai et al., 2010c, emphasis from this reference). Substantively, this means that the mediation effect is small, but statistically significant, however not robust to even small correlations between the mediator and outcome induced by an omitted variable. While from this perspective the mediation effect is weakly robust, such an omitted variable would not account for a significant increase in explained variance. With the rise of access to PrEP without a doctor's visit where the importance of frequent STI testing could be emphasized, it is possible that STI testing and STI diagnosis are correlated via an omitted variable. We also investigate the impact of such a variable and visualize the results in Figure 20.14. In this case the value of ρ is 0.210 and $\widetilde{R}^2_M \cdot \widetilde{R}^2_Y = 0.020$, i.e., this mediation effect is more robust than that of promiscuity.

7. CONCLUSION

In this chapter we provided a review of the potential outcome model and directed acyclic graphs in the service of discussing causal mediation analysis. Via the examples of the Merton Award–winning papers and the empirical example of our own, we showed that serious engagement with this body of work allows for more careful consideration of when causal claims are warranted, without the detriment to theoretical depth. Recent literature in formal mediation analysis has been particularly fruitful in clarifying previously unrecognized assumptions that underlay older methods. While this is an important methodological advancement, we argued that analytical sociologists should applaud and welcome these developments on theoretical grounds. The framework and statistical software developed by Imai and colleagues allows for analysis to elaborate on mechanisms along a causal chain. Formal mediation analysis is a statistical approach to mechanisms and, therefore, offers a different perspective to agent-based

King Makovi and Christopher Winship - 9781789906851

simulations. We believe however, that their goal is the same: to elaborate on the "black box" of regression analysis and offer a better dissection of social data.

ACKNOWLEDGEMENTS

We thank Felix Elwert, Filip Karan, Steven Morgan and Xiang Zhou for their constructive comments that improved the manuscript. We also thank Gianluca Manzo, the editor of this volume, for providing us with invaluable suggestions to revise and focus the manuscript.

REFERENCES

Aksoy, Ozan and Diego Gambetta (2016), "Behind the veil: The strategic use of religious garb", *European Sociological Review*, **32**: 792–806.

Andrews, Ryan M. and Vanessa Didelez (2020), "Insights into the 'cross-world' independence assumption of causal mediation analysis", arXiv, 2003: 10341.

Angrist, Joshua D. (1998), "Estimating the labor market impact on voluntary military service using social security data on military applicants", *Econometrica*, **66**: 249–88.

Angrist, Joshua D. and Jorn-Steffen Pischke (2008), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, NJ: Princeton University Press.

Baron, Reuben M. and David A. Kenny (1986), "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations", *Journal of Personality and Social Psychology*, **51**: 1173–82.

Bearman, Peter (2012), "On analytical sociology", *Sociologica Italian Journal of Sociology online*, **6**: 0-0, FSES Working Papers 500, Faculty of Economics and Social Sciences, University of Freiburg, Fribourg, Switzerland.

Bollen, Kenneth A. and Judea Pearl (2013), "Eight myths about causality and structural equation models", in Rina Dechter, Hector Geffner and Joseph Y. Halpern (eds), *Handbook of Causal Analysis for Social Research*, pp. 301–28, Dordrecht: Springer.

Breen, Richard (2018), "Some methodological problems in the study of multigenerational mobility", *European Sociological Review*, **34**: 603–11.

Calabrese, Sarah K., Kristen Underhill and Kenneth H. Mayer (2017), "HIV preexposure prophylaxis and condomless sex: Disentangling personal values from public health priorities", *American Journal of Public Health*, **107**: 1572–6.

Chen, Yea-Hung, John Guigayoma, Willi McFarland, Jonathan M. Snowden and Henry F. Raymond (2019), "Increases in pre-exposure prophylaxis use and decreases in condom use: Behavioral patterns among HIV-negative San Francisco men who have sex with men, 2004–2017", *Aids and Behavior*, **23**: 1841–5.

Cowan, Sarah K. (2014), "Secrets and misperceptions: The creation of self-fulfilling illusions", *Sociological Science*, **1**: 466–92.

Coyer, Liza, Ward van Bilsen, Janneke Bil, Udi Davidovich, Elske Hoornenborg, Maria Prins and Amy Matser (2018), "Pre-exposure prophylaxis among men who have sex with men in the Amsterdam Cohort Studies: Use, eligibility, and intention to use", *PloS one*, **13** :e0205663.

Eaton, Lisa A. and Seth C. Kalichman (2007), "Risk compensation in HIV prevention: Implications for vaccines, microbicides, and other biomedical HIV prevention technologies", *Current HIV/AIDS Reports*, **4**: 165–72.

Eckel, Catherine C. and Philip J. Grossman (2002), "Sex differences and statistical stereotyping in attitudes toward financial risk", *Evolution and Human Behavior*, **23**: 281–95.

Elwert, Felix and Christopher Winship (2010), "Effect heterogeneity and bias in main-effects-only regression models", in Rina Dechter, Hector Geffner and Joseph Y. Halpern (eds), *Heuristics, Probability and Causality: A Tribute to Judea Pearl*, pp. 327–36, London: College Publications.

Elwert, Felix and Christopher Winship (2014), "Endogenous selection bias: The problem of conditioning on a collider variable", *Annual Review of Sociology*, **40**: 31–53.

Frangakis, Constantine E. and Donald B. Rubin (2002), "Principal stratification in causal inference", *Biometrics*, **58**: 21–9.

Goldin, Claudia and Cecilia Rouse (2000), "Orchestrating impartiality: The impact of 'blind' auditions on female musicians", *American Economic Review*, **90**: 715–41.

Grant, Robert M. et al. (2010), "Preexposure chemoprophylaxis for HIV prevention in men who have sex with men", *New England Journal of Medicine*, **363**: 2587–99.

Grant, Robert M. et al. (2014), "Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study", *The Lancet Infectious Diseases*, **14**: 820–9.

King Makovi and Christopher Winship - 9781789906851

Harawaa, Nina T. et al. (2017), "Serious concerns regarding a meta-analysis of preexposure prophylaxis use and STI acquisition", *AIDS*, **31**: 739–40.
Hedström, Peter (2005), *Dissecting the Social—On the Principles of Analytical Sociology*, Cambridge: Cambridge University Press.
Hedström, Peter and Petri Yilikoski (2014), "Analytical Sociology and Rational-Choice Theory", in Gianluca Manzo (ed.), *Analytical Sociology Actions and Networks*, Chapter 2, pp. 57–70, Chichester: Wiley.
Hogben, Matthew and Nicole Liddon (2008), "Disinhibition and risk compensation: Scope, definitions, and perspective", *Sexually Transmitted Diseases*, **35**: 1009e1010.
Huber, Martin (2019), "A review of causal mediation analysis for assessing direct and indirect treatment effects" FSES Working Papers 500, Faculty of Economics and Social Sciences, University of Freiburg, Fribourg, Switzerland.
Imai, Kosuke, Luke Keele and Dustin Tingley (2010a), "A general approach to causal mediation analysis", *Psychological Methods*, **15**: 309–34.
Imai, Kosuke, Luke Keele, Dustin Tingley and Teppei Yamamoto (2010c), "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects", in Hrishikesh D. Vinod (ed.), *Advances in Social Science Research Using R*, pp. 129–54, New York: Springer.
Imai, Kosuke, Luke Keele, Dustin Tingley and Teppei Yamamoto (2011), "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies", *The American Political Science Review*, **105**: 765–89.
Imai, Kosuke, Luke Keele and Teppei Yamamoto (2010b), "Identification, inference and sensitivity analysis for causal mediation effects", *Statistical Science*, **25**: 51–71.
Knight, Carly R. and Christopher Winship (2013), "The Causal Implications of Mechanistic Thinking: Identification Using Directed Acyclic Graphs (DAGs)", in Stephen L. Morgan (ed.), *Handbook of Causal Analysis for Social Research*, pp. 275–99. Dordrecht: Springer.
Manzo, Gianluca (2014), "Data, Generative Models, and Mechanisms: More on the Principles of Analytical Sociology", in Gianluca Manzo (ed.), *Analytical Sociology Actions and Networks*, Chapter 1, pp. 4–52, Chichester: Wiley.
Manzo, Gianluca (2022), *Agent-Based Models and Causal Inference*, Wiley Series in Computational and Quantitative Social Science, Chichester, UK: Wiley.
Morgan, Stephen L. and Christopher Winship (2015), *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (Analytical Methods for Social Research), Cambridge, UK: Cambridge University Press.
Newcomb, Michael E., Kevin Moran, Brian A. Feinstein, Emily Forscher and Brian Mustanski (2018), "Pre-exposure prophylaxis (PrEP) use and condomless anal sex: Evidence of risk compensation in a cohort of young men who have sex with men", *Journal of Acquired Immune Deficiency Syndromes*, **77**: 358–64.
Neyman, Jerzy ([1923] 1990), "On the application of probability theory to agricultural experiments: Essay on principles", *Statistical Science*, **5**: 465–80.
Neyman, Jerzy, K. Iwaszkiewicz and St. Kolodziejczyk (1935), "Statistical problems in agricultural experimentation", *Supplement to the Journal of the Royal Statistical Society*, **2**: 107–80.
Nguyen, Trang Quynh, Ian Schmid and Elizabeth A. Stuart (2020), "Clarifying causal mediation analysis for the applied researcher: Defining effects based on what we want to learn", *Psychological Methods* (forthcoming).
Pearl, Judea (2000), *Causality: Models, Reasoning, and Inference*, Cambridge, UK: Cambridge University Press.
Pearl, Judea (2009), *Causality: Models, Reasoning, and Inference*, 2nd ed., Cambridge, UK: Cambridge University Press.
Pearl, Judea (2014), "Interpretation and identification of causal mediation", *Psychological Methods*, **19**: 459–81.
Rohlfing, Ingo and Christina Isabel Zuber (2019), "Check your truth conditions! Clarifying the relationship between theories of causation and social science methods for causal inference", *Sociological Methods & Research*, **1**: 1–37.
Rubin, Donald B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies", *Journal of Educational Psychology*, **66**: 688–701.
Rubin, Donald B. (1977), "Assignment to treatment group on the basis of a covariate", *Journal of Educational Statistics*, **2**: 1–26.
Rubin, Donald B. (1978), "Bayesian inference for causal effects: The role of randomization", *Annals of Statistics*, **6**: 34–58.
Rubin, Donald B. (1980), "Bias reduction using mahalanobis-metric matching", *Biometrics*, **36**: 293–8.
Rubin, Donald B. (1981), "Estimation in parallel randomized experiments", *Journal of Educational Statistics*, **6**: 377–400.
Rubin, Donald B. (1986), "Which ifs have causal answers (Comment on 'Statistics and Causal Inference' by Paul W. Holland)", *Journal of the American Statistical Association*, **81**: 961–2.
Rubin, Donald B. (1990), "Formal modes of statistical inference for causal effects", *Journal of Statistical Planning and Inference*, **25**: 279–92.
Runhardt, Rosa W. (2015), "Evidence for causal mechanisms in social science: Recommendations from Woodward's manipulability theory of causation", *Philosophy of Science*, **82**: 1296–307.
Russo, Federica (2009), *Causality and Causal Modelling in the Social Sciences*, Dordrecht: Springer.

Small, Mario (2017), *Someone to Talk to*, New York: Oxford University Press.
Staiger, Douglas and James H. Stock (1997), "Instrumental variables regression with weak instruments", *Econometrica*, **3**: 557–86.
Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele and Kosuke Imai (2014), "Mediation: R package for causal mediation analysis", *Journal of Statistical Software*, **59**: 1–38.
VanderWeele, Tyler (2015), *Explanation in Causal Inference: Methods for Mediation and Interaction*, New York: Oxford University Press.
Wodtke, Geoffrey T. and Xiang Zhou (2020), "Effect decomposition in the presence of treatment-induced confounding: A regression-with-residuals approach", *Epidemiology*, **31**: 369–75.
Ylikoski, Petri (2017), "Social Mechanisms", in Stuart Glennan and Phyllis Illari (eds), *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, pp. 401–12, London and New York: Routledge.
Zhou, Xiang and Geoffrey T. Wodtke (2020), "Residual balancing: A method of constructing weights for marginal structural models", *Political Analysis*, **38**: 1–20.

King Makovi and Christopher Winship - 9781789906851

21. Experiments

Giuseppe A. Veltri

The use of experiments in the social sciences has a long history. Some disciplines employ this methodology more often than others, for example psychology, but since the end of the last century, their application has become widespread in other domains, like economics, political science and sociology. One chapter is not sufficient to cover the ground of the multiple shape and forms that experiments can take, and they can vary in their settings and in their design, the latter includes both the formal elements of groups allocations, counterbalancing and other features, but it also consists of the type of experimental tasks. In this chapter I will discuss the main common methodological principles of experimental design to all these variants and, later, I will briefly discuss online experiments as population-based survey experiments because I consider this form the most promising in the context of sociological research.

However, here is a brief summary of the most common different types of experiments in terms of contexts:

- *Lab experiments*. These are almost the stereotypical experiments that anyone can summon in their mind. The main features of this type of experiment, mostly used in psychology first and in experimental economics later, are the high degree of control retained over participants and the experimental task(s) with the aim of maximising internal validity at the expense of external validity. I will discuss this trade-off later in the chapter. More and more, laboratory experiments in the domain of psychology are also relying on physiological measures such as eye tracking and electric skin conductance.
- *Field experiments*. Randomised studies that are conducted in real-world settings are often called 'field experiments'. The problem with the term is that the word-fields-refers to the setting, but the field is just one aspect of an experiment. One should identify several criteria when referring to a field experiment: whether the treatment used in the study resembles the intervention of interest in the world, whether the participants resemble the actors who ordinarily encounter these interventions, whether the context within the subjects receive the treatment resembles the context of interest, and whether the outcome measures resemble the actual outcomes of theoretical or practical interest. The four above dimensions determine the degree of 'fieldness'. Field experiments are, to a certain degree, the inverse of laboratory experiments in terms of internal and external validity trade-offs. In other words, these allow less control over the participants and the experimental task while they gain a higher degree of realism and external validity.
- *Natural experiments*. Natural experiments in the social sciences involve a range of interventions. Random or as-if random treatment assignment may stem from various sources, including a procedure specifically designed to randomise, such as a lottery; the non-systematic implementation of certain interventions; and the arbitrary division of units by jurisdictional borders. The plausibility that assignment is indeed as-if random—considered here to be one of the definitional criteria for this type of study—varies greatly. By 'standard' natural experiments, it is meant various natural experiments in

Giuseppe A. Veltri - 9781789906851

which units are assigned at random or as-if at random to categories of the independent variable—that is, to treatment and control groups. In fact, this is an umbrella catch-all category that encompasses many different types of designs. Yet, these studies are usefully distinguished from two more specific types of natural experiments: the regression-discontinuity and instrumental-variables designs. The discovery of natural experiments is thus as much art as science: there appears to be no algorithm for the generation of convincing natural experiments, and analysts are challenged to think carefully about whether sources of natural experiments discovered in one context are applicable to other settings. It is important to note that the emphasis on discovering natural experiments is not without potential problems. For some observers, the idea of scouring the landscape for instances of as-if random assignment threatens to substantially narrow the scope of substantive inquiry. Rather than pose 'questions in search of variations', some observers suggest that researchers using natural experiments tend to focus on 'variations in search of questions'. According to these critics, the recent focus among some social scientists on identifying natural experiments divorces empirical work from theory, and thus leads to the estimation of causal parameters that are not theoretically relevant (Deaton, 2009).

- *Online experiments*. Online experiments are also an umbrella term because they entail different types of experiments. The online can be a 'field' and, therefore, we can find field experiments that are in fact online experiments (see for example, Veltri et al., 2020). In theory, natural experiments can also occur online. Several terms are used synonymously for internet-based experiments: web experiment, online experiment, web-based experiment, world wide web (WWW) experiment, and internet experiment. Here we will use the term 'online experiment'. However, more in general, we can define online experiments as experiments that take place by means of computer-mediated communication tools that are online, and in which participants perform (experimental) tasks from an ICT device that is connected to the internet. Because many laboratory experiments are conducted on computers anyway, often there is little different in the experiment tasks between online and lab experiments; the same computer-mediated task can also be used in the laboratory. In distributed web experimenting, local collaborators recruit and assist participants who all log onto the same internet-based experiment (Reips, 2000). I will discuss in more detail online experiments because their latest materialisation, population-based survey experiments, are particularly interesting to the sociologist.

More recently, the use in policy evaluation and behavioural study of the randomised controlled is becoming more and more popular (Solomon et al., 2009). A randomized control trial (RCT) is a study in which people are allocated 'at random' to receive one of several interventions. Typically, RCTs seek to measure and compare different outcomes that are present or absent after the participants receive the interventions. Because the outcomes are quantified (or measured), RCTs are regarded as 'quantitative' studies. RCTs are used to compare two or more interventions and are therefore considered to be 'comparative' studies. Usually, one of the interventions is regarded as a standard of comparison or 'control'.

An RCT is an experiment whereby study participants are assigned by chance, following a pre-test, to at least two conditions: an experimental treatment or intervention, and a control intervention used for purposes of comparison on outcomes. Research study participants are offered their assigned intervention and are expected to participate in their assigned treatments

Giuseppe A. Veltri - 9781789906851

to at least a nominal extent, and then are compared by group on some outcome(s) that are theoretically linked to the interventions being compared. Random assignment enables direct, causal attribution of any difference in outcome to the intervention assignment and systematically rules out rival, plausible explanations other than the experimental versus control condition, thus enabling cause and effect determinations. RCTs are considered to be the 'gold standard' of evidence when determining the effectiveness of policy and practice interventions. With increasing emphasis on providing empirical evidence on the effectiveness of psychosocial interventions by various professions, the use of RCTs is growing in prominence.

In the author's view, RCT are slightly different in terminology when applied in clinical settings compared to their use in psychology and other social science disciplines. The different forms of RCT are: parallel group—each participant is randomly assigned to a group, and all the participants in the group receive (or do not receive) an intervention; crossover—over time, each participant receives (or does not receive) an intervention in a random sequence; cluster—pre-existing groups of participants (such as villages and schools) are randomly selected to receive (or not receive) an intervention; factorial—each participant is randomly assigned to a group that receives a particular combination of interventions or non-interventions. All these variants correspond to design solutions present in randomised experiments; for example, parallel-group RCT in clinical settings are equivalent to between-subjects randomized experiments in psychology and the social sciences. The same is true for crossover RCTs that are equivalent to within-subjects experiments.

After this brief overview of these common forms of experiments in the social sciences, in this chapter I want to provide the basic 'grammar' of experimental design for analytical sociologists who are not familiar with this method in details (for a previous brief exposition of experiments for analytical sociologists, see Bohnet (2009). First, I provide a brief overview of where experiments are situated in the broader epistemological discussion about the determination of causality in the social sciences. Second, I introduce some basic elements of design that are shared by almost all types of experiments (with the exclusion of a quasi-experiment that I will not cover in this chapter for the sake of brevity). Last, I will discuss in more detail online experiments and their population-based version.

1. APPROACHES TO CAUSALITY

Before moving on to describe the basic features of experiments shared by all their different forms, it is useful to briefly discuss the issue of causation in the social sciences. On this topic, literally hundreds of books and articles have been written and the possibility of causal explained social phenomena is fiercely debated. There are those who do not believe it to be possible (e.g. Martin, 2011) and others who do (e.g. Woodward, 2000). It is not possible to fully recount such a debate, and it is not among the aims of this chapter to do so, but we need to be clear about one distinction between forms of causality in order to understand the difference between experiments and other methods.

Broadly speaking, in the social sciences we have two separate forms of understanding causality: *regularity theories of causation* and *manipulability theories of causation*. Regularity theories are based on a long philosophical tradition that can be traced to David Hume, who defines causality as determined by temporal priority (the cause must precede the effect), spatiotemporal contiguity (cause and effect are near in time and space), and constant conjunction

Giuseppe A. Veltri - 9781789906851

(if the same circumstances are repeated, the same outcome will occur). The tradition stemming from this position on causality has focused on distinguishing between causal and noncausal relationships between variables (e.g. Yule, 1896; Simon, 1954). Starting with Hume, regularity theories of causation have been concerned with the full (or philosophical) cause of an effect, by which is meant a set of conditions that is sufficient (or necessary and sufficient) for the effect to occur. The further elaboration of probabilistic causation in the regularity theories led to the development of concepts like spurious relationships[1] and mediation effects between variables. Granger causation, a concept based on prediction, and structural equation models, use this type of reasoning to distinguish between causal and noncausal empirical relationships. One of the implications of the regularity approach to causation is the need to specify all the variables and the functional relationship between these and the outcome. Analysis done with survey data or other sources of quantitative information use this approach to causality.

Experiments are, however, grounded on a different tradition of understanding causality: manipulability theories. This approach views causes as variables that can be manipulated, with an outcome that depends on the state of the manipulated variable. Rather than accounting for the relationships between several variables, including mediating ones, so as to reach an outcome, the goal is to examine the effect of a particular variable. To some, it is not impossible to reconcile these two approaches, regularity and manipulability (Sobel, 1995); others argue in favour or such combination (see Chapter 16 in this book). Manipulability theories require the causal relation to sustain a counterfactual conditional statement (e.g. eating the poison that caused Marc to die; had he not eaten it, he would not have died). There are problems with this approach to causality too, particularly in terms of counterfactual theories (see Martin, 2011).

There is a third, more recent approach to understand causality, and this is the *generative approach* (Epstein, 1999). The connection between individuals' actions and their collective consequences would be transparent if one could simply sum over individuals' intentions or behavior to generate expected population-level attributes. The problem is that nearly all human behavior is interdependent; individuals' actions are contingent on the past, present, and predicted future behavior of others. A key feature of agent-based modeling is that it explicitly links micro and macro levels of analysis. Sociology has a longstanding interest in the relationship between individuals' motivations and decisions and large-scale patterns of social organisation and change. The micro-macro problem concerns how to explicitly account for the ways in which actions of individuals give rise to social organisation and dynamics, rather than assuming that macro-level phenomena are simply aggregates of individual characteristics and behavior. For a review of the potential and limitations of agent-based models, see Manzo (2014a) and Chapter 20 in this book.

Experiments are therefore situated in the manipulable understanding of causality, and their design is characterised by three main aspects: (1) manipulation of one or more independent variables; (2) use of controls such as randomly assigning participants to treatment or experimental or control groups; (3) careful observation and measurement of one or more

[1] From Wikipedia: In statistics, a spurious relationship or spurious correlation is a mathematical relationship in which two or more events or variables are not causally related to each other, yet it may be wrongly inferred that they are, due to either coincidence or the presence of a certain third, unseen factor (referred to as a 'common response variable', 'confounding factor' or 'lurking variable')."

Giuseppe A. Veltri - 9781789906851

dependent variables. An experimental design identifies the independent, dependent, and nuisance variables and indicates the way in which the randomisation and statistical aspects of an experiment are to be carried out. The primary goal of an experimental design is to establish a causal connection between the independent and dependent variables. A secondary goal is to extract the maximum amount of information with the minimum expenditure of resources. Interestingly enough, the manipulability understanding of causality finds a receptive audience in the analytical sociology engaged with the generative approach. The use of experiments as input for parameters calibration in multi-agents models at the micro level has been called for by many scholars of this field (see, for example, Manzo, 2014b).

Next, I consider two crucial concepts for understanding the logic of experiments that are particularly relevant for their web version: random assignment and the notion of validity.

1.1 Random Assignment

One crucial concept in experimental design is the notion of *random assignment* (Montgomery, 2009). A random assignment in experimental research has three important functions: (1) it distributes individual characteristics of participants across the treatment and control group so that they do not systematically affect the outcome of the experiment; (2) it helps ensure that error effects, which refer to the effect on the outcome variable not attributable to the manipulated variable, are statistically independent; (3) it creates a group of respondents who are, at the time of such division, probabilistically similar on average. If the research has all the features of an experiment except random assignment, it is called a *quasi-experiment*. Unfortunately, the interpretation of quasi-experiments is often ambiguous. In the absence of random assignment, it is difficult to rule out all variables other than the independent variable as explanations for an observed result. Moreover, quasi-experiments may not have a control group, which further complicates their validity.

In order to explain the role of a random assignment, let us see an example of an experiment in which our treatment group consists of subjects being exposed a given set of information in the form of a pop-up message during the use of a given online platform. The control group consists of subjects interacting with the same platform without being shown the pop-up message (see a real example of such experiment related to online gambling in Folkvord et al., 2019).

The logic underlying randomised experiment is often explained in terms of a notational system that has its origins in Neyman (1923) and Rubin (1974) and that I present here in a summarised and simplified version. The formalised notation is as follows:

We have observation on N units indexed by $I = 1, \ldots, N$

We use the binary indicator D_i to indicate whether the unit *i* was treated or not treated as a dummy variable with $D_i = 0$ if unit *i* is not treated (not shown the pop-up), and D*i* = 1 if unit *i* is treated (was shown the pop-up)

> Y_i is the observed outcome (provided or not personal information) for each population unit *i*
>
> $Y_i = Y_{1i}$ if $D_i = 1$ (unit *i* is treated)
>
> $Y_i = Y_{0i}$ if $D_i = 0$ (unit *i* is not treated)
>
> Let $(Y_1, Y_0)_i$ be the two outcomes corresponding to the *i-th* population unit being treated or not treated, respectively
>
> $Y_{1i} - Y_{0i}$ = unit-level causal effect

Giuseppe A. Veltri - 9781789906851

Table 21.1 Factual and counterfactual outcomes

	Factual outcome	Counterfactual outcome
Treated (D = 1)	Y_1	Y_0
Nontreated (D = 0)	Y_0	Y_1

Y_{0i} = is the counterfactual for an individual who was shown the pop-up.
Y_{1i} = is the counterfactual for an individual who was not shown the pop-up.

If a specific member of the population is exposed to the intervention, then Y_1 is observable, while Y_0 is irreversibly unobservable on that specific member. It follows that Y_1 and Y_0 are the factual and observable outcomes for treated and nontreated, respectively. The counterfactual outcome for a member of the population who was treated is Y_0 and for one who was not treated is Y_1. Henceforth, we simplify the notation by abandoning the index *i* and simply using the suffix 1 or 0, but with the important observation that blue refers to factual outcomes and red to counterfactual ones, as shown in Table 21.1.

Now let us proceed by illustrating the standard logic identity focusing on the average treatment effect on the treated.[2]

Average Treatment on the Treated (ATT) = $E\{Y_1 - Y_0 \mid D = 1\} = E\{Y_1 \mid D = 1\} - E\{Y_0 \mid D = 1\}$

Evidently the second term of the identity is unknown and unobservable as it represents the counterfactual for the treated; it means a hypothetical subject who saw the pop-up would have done so had they not been shown the pop-up. Even more concretely, if we have Marc being shown the pop-up and deciding not to provide personal information, then the second term asks what Marc would have done had he not been shown the pop-up. Would he have provided the personal information or not?

Let us now consider the term $E\{(Y_1 \mid D = 1) - E\{Y_0 \mid D = 0\}$, which represents the factually observed difference in means between treated and nontreated that is observed and known; then we add and subtract to it the term $E\{Y_0 \mid D = 1\}$, which is the counterfactual for the nontreated. (Laura was not shown the pop-up. What would she have done had she been shown it? Again, this is unknown and unobservable.)

We, thus, get: $\mathbf{E\{(Y1 \mid D = 1) - E\{Y0 \mid D = 0\}} = \underline{[E\{Y_1 \text{ I } D = 1\} - E\{Y_0 \text{ I } D = 1\}]} + [\{E\{Y_0 \text{ I } D = 1\} - E\{Y_0 \mid D = 0\}]$; this means that **observed means differences** = Average Treatment on the Treated (ATT) + *Selection Bias*. So, the ATT and the selection bias cannot be disentangled and simply calculating the difference in observed means will not suffice to identify the causal effect of the treatment on the treated. The selection bias term of the identity, $[\{E\{Y_0 \text{ I } D = 1\} - E\{Y_0 \mid D = 0\}]$, can be easily interpreted as the difference in the outcome Y that would be observed between treated and nontreated if there was no treatment, and depends on pre-existing differences between the two groups. That is, this term captures outcome differences between the treated and nontreated

[2] This is the causal effect of major interest from a policy perspective; the other important effect is the average treatment effect (ATE) defined as E {Y1 – Y0} or the average effect that would result from having all population members (or none at all) take part in the programme. ATE perhaps is not a crucial parameter of interest for impact evaluation. It is the most relevant parameter when the programme under consideration is universal, in the sense that it would expose all units of the target population to treatment.

Giuseppe A. Veltri - 9781789906851

that cannot be attributed to the treatment. The first term of the selection bias is not observable. So, unless the selection bias is equal to zero or we can in some way control it (i.e. measure it), the full identity illustrated cannot be given a causal interpretation. The selection bias is a function of X characteristics (possibly unobserved) of the treated and nontreated subjects. In other words, we are saying that a vector of X variables shapes the values of the variable D, or how the groups D = 0 and D = 1 are formed. Using some additional equations, which we will not replicate here, we can show that X is correlated to Y and hence the selection bias does not vanish, since the two groups (treated and nontreated) are not equivalent with regard to Xs. In other words, there is a selection bias when (1) being treated or not depends on individual characteristics X known to affect also the outcome variable; and (2) such characteristics are unequally distributed between the treated and the nontreated.

Under the condition of a randomised experiment, the selection bias vanishes and then we can basically recover the causal effect as a simple difference in means. Under randomisation, the dummy variable D depends on the toss of a coin and is not systematically related to either Y or Xs in anyway, hence the selection bias of the previous identity entirely vanishes (equal to zero). In effect, if we randomise subjects either to be exposed to a pop-up or not, their unobserved characteristics would no longer be related to how they browse a website and the related changes of coming or not across the pop-up. Well-performed randomisation ensures that the treated and nontreated groups are balanced with respect to these unobserved characteristics.

Under the condition of randomisation, then, the treatment effect is estimated as a difference between two means. Statistically, such a difference in mean can be obtained (if it exists) by running a regression analysis where the dependent variable is our outcome Y (in the exemplification we used the provision of personal information) and the independent variable is a dummy equal to one if the subject has seen the pop-up, and zero otherwise. Let 'y' be, for instance, 'provision of personal information' measured with respect to whether subjects saw a pop-up (D = 1) or not (D = 0). Hence, the general formula for the y is:

$$y = a + b\,D_1 + \boldsymbol{\varepsilon}.$$

The term $\boldsymbol{\varepsilon}$ includes all the observable and unobservable variables that impact on Y. What is important is that under randomisation the expected value of $\boldsymbol{\varepsilon}$ under treatment (in technical notation, $E[\boldsymbol{\varepsilon}|\, D_1]$) is equal to zero (orthogonality condition). A simple ordinary least square (OLS) regression is capable of identifying 'b', since the exogeneity condition is met. The coefficient 'b' is precisely the treatment effect. In fact, for those that did not receive the treatment, the expected value for y is

$$\mathrm{E}[y \mid D_1 = 0] = y_0 = a.$$

while for the treated group, we have

$$\mathrm{E}[y \mid D_1 = 1] = y_1 = a + b.$$

So, by definition

$$b = y_1 - y_0.$$

Giuseppe A. Veltri - 9781789906851

This corresponds to the difference in the mean outcome between the treated and untreated groups, i.e. the treatment effect. The coefficient 'b' is also what we can consistently estimate with OLS.

When subjects are randomly assigned to conditions, any unknown and unknowable factors are distributed equally and thus (1) *the causal effects of those unknown and unknowable factors will be equated across experimental and control groups*; and (2) *factors that are the same can never cause a difference*. Thus, any differences can be reasonably attributed to variation from the independent variable. Although Fisher's principle of random assignment is widely recognized as the panacea, others still believe the method is grievously flawed. One notorious critique is one by Bernard Cohen (1989, 1997) who argues that as the set of unknown or unknowable causes grows larger, the probability that the experimental and control groups will differ on at least one of these causal factors approaches 1.0. Furthermore, Cohen notes that random assignment only equates experimental and control groups with infinite sample sizes and, of course, real experiments always employ finite samples. He therefore concludes that there is always some unknown causal factor operating in any experiment. His analysis centers on (1) the number of potential alternative causes; and (2) the number of subjects in the experimental and control groups—claiming that the former is too large and the latter is too small for random assignment to properly operate. However, in a mathematical sense, Cohen (1997) overestimates both the number of alternative causes that *do* exist and the sample size that *is* required for random assignment to work properly. In terms of sample size, Cohen (1997) ignores the straightforward statistical relation between sample size and the nature of alternative causes. That relationship is described by the *law of large numbers*. This law dictates that as the size of the experimental and control groups increases, the average value of any factor differentiating those groups approaches a common value (i.e. the population value for that factor). This law also has implications for the *number* of alternative causes. The law implies that as group size increases and alternative causal factors become equated, the number of *differentiating possible causes will diminish at an exponential rate*. As this occurs, the probability that the experimental group differs from the control group, on any single factor, approaches zero.

There are additional logical and empirical grounds that further salvage the method. Logically, it is important to distinguish factors that could make a difference from factors that do make a difference. Not all of the factors that make experimental groups different from control groups are relevant to the dependent variable and, therefore, not all factors must necessarily be equated. Many differences simply do not matter. Scientists specifically control for *theoretically relevant* factors in a given study, and in true experiments they also control for *possibly relevant* factors using random assignment. Cohen heavily emphasizes those factors that are possibly relevant. However, given that experimental research selectively focuses on specific theoretical problems, most of the infinite set of possibly relevant differences probably will be irrelevant.

1.2 Validity

Random assignment enables the researcher to formulate the appropriate comparisons, but random assignment alone does not ensure that the comparison will speak convincingly to the original causal question. A randomised experiment obviously presents an ideal opportunity for having (1) the data generating process (DGP) of the variable of intervention subject to exogenous variation; (2) the participants to the study representative of the population object of the intervention; and (3) the behaviour measured in the lab (on average) identical to the

Giuseppe A. Veltri - 9781789906851

variable of interests that we want to study with respect to our theory and/or the underlying real environment phenomenon of interest. The main strategic challenge is that experiments may remain a very 'local' type of evidence, and this is problematic if our goal is to support policymakers who are interested in general results. In fact, all the literature on experimental validity, as well as on internal and external validity, originated in the theory that Campbell and his associates developed on how to move from local evidence to generalisation. The problem was framed as one of extension to other populations and settings (external validity) and of representation and measurement (construct validity). As noted (McGraw and Hoekstra, 2010; Shadish et al., 2002), in this effort, Campbell early on abandoned the single binary distinction between internal and external validity; first, Cook and Campbell (1979) extended the notion of internal validity and, later (see Shadish et al., 2002), he incorporated clarifications advanced by Cronbach with his UTOS (units, treatments, observation operations, settings) framework (Campbell, 1957); validity was divided into four concepts: causal validity, construct validity, statistical validity, and external validity. In addition is ecological validity, which, as we will see, partially overlaps with both construct validity and external validity, although it has its own distinct tradition in psychology, starting with Brunswik's (1943) suggestion that we should study the participant's setting, it is then further developed, particularly from an ecological perspective. In the sections that follow, I will describe the key aspects of different dimensions of validity before briefly illustrating the main threats to validity. I will conclude with an analysis of trade-offs and of the different traditions and research purposes that explain the choices made with respect to them.

1.3 Internal Validity

If validity in general is the approximate truth of the inference or knowledge claim, then internal validity is the approximate truth of the inference or knowledge claim made with respect to a particular target population. Following the same general definition, external validity would be the approximate truth of the inference or knowledge claim for observations beyond the target population studied.

Originally, Campbell (1957) considered an experiment internally valid if the experimenter finds a significant difference between the treatment and control conditions. These differences are then assumed to provide a meaningful reflection of the causal processes at play. As long as no reason exists to assume that some extraneous mediating factor systematically influenced subjects' responses, observers can attribute changes in the dependent variable to systematic manipulations across the independent variables. From this perspective, internal validity is enhanced by experiments that are well designed, carefully controlled, and meticulously measured, so that alternative explanations for the phenomena under consideration can be excluded. Shadish et al. (2002) are careful to note that internally valid findings remain discrete to the specific experimental context in which they are explored. As we have seen, Campbell and associates divided internal validity into three separate concepts: causal validity, construct validity, and statistical validity.

Causal validity determines whether the relationships found by the researcher within the target population analysed are causal. Establishing causal validity for the relationship between X and Y would mean establishing that changes in X cause changes in Y. Very often this is what analysts have in mind when they consider internal validity, whereas establishing a causal relation does not exhaust the criteria for assessing the internal validity of an experiment, if we recall that the goals are those of extension from local to general knowledge and of

Giuseppe A. Veltri - 9781789906851

representation. We can possibly demonstrate that X causes variations in Y and that there are no other factors influencing them and yet score poorly in internal validity if the inferences we draw are flawed; for instance, by the fact that the variables we measured do not represent well the variable either in theory or in practice, which smoothly leads us to consider the next dimension: construct validity.

Construct validity has, first and foremost, to do with how valid the inferences of the data are for the theory (or constructs) the researcher is evaluating. In experimental research, the question is whether the design of the experiment is such that the variables investigated are closely equivalent to the variables the theory is concerned with. Are those things that the theory holds constant held constant in the experiment? Are the choices before the subjects the same as the choices assumed in the theory? In other words, is there a close match between what the theory (the real world) is about and what is happening in the manipulated DGP? As the theory can be assumed to deal with what happens in the real world, then these sentences can to some extent be rephrased in terms of the adherence between the experiment and what happens in practice; for instance, are the choices before the subjects the same (or similar enough) as the choices they face in the real world? In this respect, construct validity and ecological validity overlap and the former is not immune from consideration in the trade-off between mundane and experimental realism (see below). So, although construct validity is a dimension of internal validity, construct validity is also about generalisation and reflects the attempt to move from the local to the general by way of ensuring valid representation. Results from experiments with high construct validity can help us answer more general questions than those without construct validity.

Statistical validity, in simple terms, depends on the significance of the statistical relation estimated for the variables of interest and on the size of such a relation. Given the assumption made, and other aspects of the design and of the DGP, one must ask whether estimates are efficient, accurate, significant, and sizeable, and also whether the dataset is representative of the target population. Usually, other aspects of validity are considered more challenging and given more attention than statistical validity, as the latter is, so to speak, 'commoditised' given the increasing capacities of statistical software packages. Yet, there are questions that statistical packages do not solve and on which the research must reflect. What in general are the implications of different levels of statistical significance? What if we find a relationship that is just on the edge of the 5 per cent level? Also important is the extent to which the assumptions about variables distribution are supported and whether errors are correctly estimated.

Mundane versus experimental realism: As argued by Aronson et al. (1990), despite efforts to prove otherwise, internal validity remains intrinsically tied to experimental, as opposed to mundane, realism. Those who are sceptical about the experimental method often point to the artificial nature of laboratory settings as a main weakness and an obstacle to generalisation. The critique holds that tasks presented to subjects offer a poor analogy to the real-world experiences that individuals confront. Strong proponents of the experimental method, especially in economic game-theoretic experiments, but also to some extent in psychology, argue on the contrary that stylised and stripped-down settings are needed to carefully operationalise and measure the variables of interest and then, through multiple tests on numerous populations, to begin to define the conditions under which generality might obtain. The reason it becomes so critical to uncover these mechanisms is because unless an investigator knows the underlying principles operating in a given dynamic, it will simply prove impossible to ascertain which aspect of behaviour is causing which effect within the context of real-world settings, where many other variables and

Giuseppe A. Veltri - 9781789906851

interactions occur simultaneously. In this tradition, it does not matter whether the experimental environment does not overtly mimic the real-world setting, as long as the subject experiences the relevant forces that the investigator seeks to elicit. A more balanced view is to recognise that, if experimental subjects become psychologically disengaged in the process, they confront and do not pay attention to performing the tasks, and this clearly undermines internal validity and the possibility of extending the findings. If subjects approach a task with scepticism or detachment, then genuine responses fade and strategic incentives come to the fore. This raises the possibility that measures obtained do not accurately reflect the process being manipulated but, rather, manifest a different underlying construct altogether. This is a clear threat to construct validity. Hence, the motivation of the subjects and the internal experience of the experiment for them should ensure engagement and, for this to occur, the settings need not necessarily reflect outside appearances. The success of the experiment depends on the subject taking the task seriously, and experimenters can foster such engagement to the degree that they can create and establish a situation that forces psychological investment on the part of subjects. The critical operative feature in such experimental designs revolves around the ability of the experimenter to create a psychological situation that realistically elicits the dynamics under consideration. So, experimental realism remains more important than mundane realism in maximising prospects for internal validity because it is more likely to elicit the critical dynamic under investigation, but more highly stylised or abstract experimental protocols can risk both internal and external validity by failing to engage subjects' attention or interest.

1.4 External Validity

External validity refers to the generalizability of findings from a study, or the extent to which conclusions can be applied across different populations or situations; too often, however, external validity is simply intended as a matter of representativeness or of the size of the sample. Privileging external validity often results from a misunderstanding that generalizability can result from a single study as long as it is large enough or broad enough; this is never true. External validity results, primarily and eventually, from replication of particular experiments across diverse populations and different settings using a variety of methods and measures. As Aronson et al. (1990) put it: 'No matter how similar or dissimilar the experimental context is to a real-life situation, it is still only one context: we cannot know how far the results will generalise to other contexts unless we carry on an integrated program of systematic replication' (p. 77). External validity is the extent to which the 'causal relationship holds over variations in persons, settings, treatments, and outcomes' (Shadish et al., 2002: 83). Hence, it covers four aspects of experimental design: (1) whether the participants resemble the actors who are ordinarily confronted with these stimuli; (2) whether the context (including the time) within which actors operate resembles the context (and time) of interest; (3) whether the stimulus used in the study resembles the stimulus of interest in the world; and (4) whether the outcome measures resemble the actual outcomes of theoretical or practical interest. So, the challenge for external validity is to extend to other subjects (i.e. representativeness *sensu stricto*) other observations, other treatments, and other settings as in the UTOS framework articulated by Cronbach (1982) for establishing valid causal inference. Finally, it is worth pointing out that, often, when critics say that an experiment does not have external validity, they just have in mind that the sample used (in the laboratory) is not randomly drawn; yet, this has to do with internal statistical validity and not with external validity. Random sampling from a target population does not

Giuseppe A. Veltri - 9781789906851

necessarily mean that a result is externally valid. External validity has to do with generalising to populations beyond the target population, so whether one has a random sample from the target population tells us nothing about the external validity for other populations for which one has not taken a random sample.

With each innovation comes the requirement to ensure that the methods reveal what they claim they do. For web research, there seem to be two primary ways to establish validity: (1) compare results from a web-based study with a laboratory-based study and (2) examine the research to see if the results follow theoretically predicted trends. Web experiments may be used to validate results from field research and from laboratory experiments, or they may be used for new investigations that could only be feasibly accomplished in this medium. Because many laboratory experiments are conducted on computers anyway, nothing is lost when an experiment is designed web-ready; it can always also be used in the laboratory. In distributed web experimenting, local collaborators recruit and assist participants who all log onto the same internet-based experiment. External validity of web experiments may be limited by their dependence on computers and networks. Also, many studies cannot be done on the web. However, where comparable, results from web and laboratory studies are often identical (Reips, 2002a).

1.5 Ecological Validity

As a starting point, ecological validity can be seen as referring to the relation between real-world phenomena and the investigation of these phenomena in experimental contexts. Yet, as for other dimension of validity, there are many different definitions (Scheidt, 1981: 225–6). The focus on ecological validity started with Brunswik's (1943) suggestion to study the participant's setting. It then continued with Bronfenbrenner's (1977) echoing of the same emphasis and with Neisser (1976) arguing that the stimulus materials should be ecologically valid. In brief, ecological validity concerns focused on the realism of three aspects: the settings, the stimuli, and the responses. Ecological validity involves maintaining the integrity of the real-life situation in the experimental context while remaining faithful to the larger social and cultural context. More intuitively ecological validity is similar to what Harrison and List (2004) refer to as the 'fieldness' of an experiment.

Ecological validity is a dimension that could be considered a subset of both external validity and construct validity. If we consider it in terms of settings (the S in Cronbach's UTOS framework) then, obviously, ecological validity is a subdimension of external validity. On the other hand, if we look at ecological validity as concerning the problems that may arise from the interaction between the aseptic (artificial) context typical of a laboratory experiment and the behaviour of subjects, which means the interaction between the design and the measurement of the variable of interest, then lack of ecological validity can actually be seen as a problem of construct validity inasmuch as the measurement generated may not validly measure the underlying construct (phenomenon) because of, for instance, subjects' disengagement with the experimental task. It is about the similarity between the environment constructed in the research and a target environment. Some experimentalists call this mundane experimental realism or contextual congruence. The experimental environment is considered ecologically valid if the methods, materials, and settings of the research are similar to the target environment.

It must be clarified, however, that maximising ecological validity may or may not enhance external validity of the results because it may not be possible to generalise the target environment. External validity can be conjectured or hypothesised based on similar studies

Giuseppe A. Veltri - 9781789906851

or assumptions about population similarities concerning any study, experimental or non-experimental, but the proof of external validity is always empirical.

1.6 Main Threats to Validity

Campbell and Stanley (1966) identified nine threats to internal validity that may lie outside the experimenter's ability to control: selection, history, maturation, repeated testing, instrumentation, regression toward the mean, mortality (attrition), experimenter bias, and selection–maturation interaction. Here, I consider only the most important of these.

Noncompliance can be active or passive and may occur in different forms; one is when those assigned to the treatment group do not receive the treatment, or when those assigned to the control group inadvertently receive the treatment. In this case, the randomly assigned groups remain comparable, but the difference in their average outcomes measures the effect of the experimental assignment rather than actually receiving the treatment. Another form of noncompliance is disengagement that is more subtle, but as serious as, active or passive forms of noncompliance. Many standard forms of analysis assume that everyone who receives treatment experiences a similar level of engagement with the protocol, but this assumption is often faulty because noncompliance rates can be nontrivial. Noncompliance raises the prospect of artificially inducing treatment values into the estimated outcome effects. If subjects are not paying attention, or are thinking about something unrelated to the task at hand, then it will remain unclear whether the manipulation actually exerted an effect.

Attrition involves the failure to measure outcomes for certain subjects (e.g. some do not report their vote preference in the follow-up) and is particularly problematic when it afflicts some experimental groups more than others. Attrition occurs when subjects drop out of an experiment or are otherwise lost to follow-up. This only poses a threat to internal validity to the extent that it occurs subsequent to random assignment (Kiesler et al., 1969). If subjects drop out prior to such an assignment, then it may constitute a threat to the external validity of the experiment (sample characteristics), but not to the internal validity of its findings. However, if subjects in one condition are dropping out of an experiment at a higher rate than those in another condition, then it may be that such attrition is in fact affected by the treatment itself.

Experimental effects (or, in experimental economics, experimenter demand effects) refer to the way in which people change their behaviour simply because they know they are being monitored. Subjects may get from the situation or directly from the experimenter cues about what constitutes appropriate behaviour and act accordingly, or against it. They are usually a potential problem only when they are positively correlated with the true experimental objectives' predictions, and we identify techniques such as nondeceptive obfuscation to minimise this correlation. Harrison and List (2004) argued that manipulation by an experimenter puts subjects on an 'artificial margin' and may lead subjects to alter their behaviour on unconstrained margins. There are different types of effects; in psychology they have been classified precisely (Howitt and Cramer, 2011: 202–6). One particularly important subclass is referred to as the 'experimenter expectant effect' when, unintentionally, the researcher convinces subjects to act or behave in a particular way in order to achieve what the experimenter wants. Another such subclass is known as 'demand characteristics', which occurs when subjects influenced by the experimental situation (that gives cues to the hypothesis) alter how they act.

A different source of effects potentially undermining validity is the so-called 'experimental cross-effect' that can occur (1) when subjects' choices in an experiment are influenced by the

Giuseppe A. Veltri - 9781789906851

manipulations they have received in previous and distinct experiments; and (2) when subjects' choices in one task of a given experiment are influenced by the manipulations they have received in a previous task of the same experiment. This threat is particularly relevant in within-subjects designs, where it can be the source of order and learning effects and increases the other experimental effects cited earlier, including social desirability and hypotheses guessing.

Restricted sample and selection bias. Restricted subject populations can also limit the degree of potential generalizability from studies, although the degree to which this problem poses a serious threat varies with the topic under investigation. Although it is generally better to have more subjects across a wider demographic range, depending on the content of the study, it may be more important to obtain more subjects, rather than explicitly diverse ones. Selection bias, in terms of non-random sampling, represents another threat to external validity that also threatens internal validity. If subjects are drawn from too restrictive a sample or an unrepresentative sample, then obviously more replication will be required to generalise the results with confidence.

1.6.1 Trade-offs between different forms of validity

Obviously, it is best to strive to maximise both internal and external validity, but this may run counter to practical and logistical constraints. Maximising internal validity may diminish the ability to extrapolate the findings to situations and populations outside those specifically studied. Privileging external validity often neglects important aspects of internal experimental control so that the true cause of reported findings remains unclear. Two principal trade-offs exist between internal and external validity. Attention to internal validity optimises the ability of an investigator to achieve confidence that changes in the dependent variables truly resulted from the manipulation of the independent variable. In other words, methodological and theoretical clarity emerge from careful and conscientious documentation of variables and measures. In contrast, concentration on external validity by expanding subject size or representativeness can increase confidence in generalizability, but only to the extent that extraneous or confounding hypotheses can be eliminated or excluded from contention. Regardless of the solutions adopted, the trade-off will always remain between the ability to control and measure carefully defined variables and testing in more realistic settings for the purpose of ecological and external validity. External validity and the capacity to extend from the local to the general remains legitimate, but it is a concern that should arise only after prior sufficient attention is given to internal validity first. As argued by Aronson et al (1990).: 'Internal validity is, of course, the more important, for if random or systematic error makes it impossible for the experimenter even to draw any conclusions from the experiment, the question of the generality of these conclusions never arises' (p. 75). A single, very large (in terms of sample) and very contextually rich and realistic experiment will never ensure external validity as such. External validity will come with replication across time, treatments, populations, and settings.

Having said this, differences in the extent to which researchers give more priority to one or the other dimension of validity remain and depend on the disciplinary tradition and/or the purpose for which experiments are carried out. Disciplines such as political science and sociology have been slow to adopt the experimental method; they are more concerned with generalisation than, for instance, psychology and experimental/behavioural economics. Whereas in political science and sociology the concerns with external validity are at times excessive, the reverse is true for psychology and experimental economics with their priority on internal validity (Veltri, 2019; Salganik, 2019). The purpose of the experiment, not always coinciding with the

Giuseppe A. Veltri - 9781789906851

disciplinary tradition, also plays a role. If the goal of the experiment is to extend theoretical models or data generation, then internal validity is the priority, especially in game theoretic experiments undertaken by experimental economist stylisation and aseptic settings, which are sought for the sake of internal validity. On the other hand, if the purpose of an experiment is to support policymaking, then external validity tends to have at least as much importance as internal validity because policymakers are interested in the potential extension of results to large and different populations and settings. And, yet again, it is important to remind ourselves that striving for external validity at the cost of internal validity will not make the evidence base of policymaking any better. We can quote again Aronson et al. (1990) again:

> Bringing the research out of the laboratory does not necessarily make it more generalisable or 'true'; it simply makes it different. The question of which method is—'artificial' laboratory experiments versus experiments conducted in the real world—will provide the more generalisable results is simply the wrong question. The generalizability of any research finding is limited. This limitation can be explicated only by systematically testing the robustness of research results across different empirical realisations of both the independent and dependent variables via systematic replication to test the extent to which different translations of abstract concepts into concrete realisations yield similar results. (p. 82)

In conclusion, the concrete purpose of the experiment will shape the needed balance between internal and external validity. Exploring a fairly universal human experience (i.e. some very basic emotional reaction through visual stimuli for instance) it is very unlikely that college students will differ from other population groups, or even from college students from other countries. In other situations, external validity will instead require having a broader sample in terms of representation of different sociodemographic segments.

1.7 Types of Design

Regardless of where your experiment takes place, in the field, in a lab, or online, there are basic principles in the composition of an experimental design. Experiments that focus on one independent variable (or manipulated variable), called 'single-factor' experiments, and those that study the role of multiple independent variables, or 'multi-factor' experiments. The latter can have several levels assumed by the independent variables. For example, a 2 x 2 multiple-factor design is a study that refers to the structure of an experiment that studies the effects of a pair of two-level independent variables. The independent variables are manipulated to create four different sets of conditions, and the researcher measures the effects of the independent variables on the dependent variable.

The advantage of factorial designs can be easily illustrated. Suppose we have two factors, A and B, each at two levels. If we denote the levels of the factors by A^-, A^+, B^-, and B^+, the effect of changing factor A is given by $A^+B^- - A^-B^-$, and the effect of changing factor B is given by $A^-B^+ - A^-B^-$. Because experimental error is present, it is desirable to take two observations, say, at each treatment combination, and estimate the effects of the factors using average responses. Thus, a total of six observations are required. If a factorial experiment had been performed, an additional treatment combination, A^+B^+, would have been taken. Now, using just *four* observations, two estimates of the A effect can be made: $A^+B^- - A^-B^-$ and $A^+B^+ - A^-B^+$. Similarly, two estimates of the B effect can be made. These two estimates of each main effect could be averaged to produce average main effects that are *just as precise* as those from the

Giuseppe A. Veltri - 9781789906851

single-factor experiment, but only four total observations are required, and we would say that the relative efficiency of the factorial design to the one-factor-at-a-time experiment is (6⁄4) = 1.5. Now suppose interaction is present. If the one-factor-at-a-time design indicated that $A^{-}B^{+}$ and $A^{+}B^{-}$ gave better responses than $A^{-}B^{-}$, a logical conclusion would be that $A^{+}B^{+}$ would be even better. However, if interaction is present, this conclusion may be seriously in error.

Factorial designs have several advantages and are more efficient than one-factor-at-a-time experiments. Furthermore, a factorial design is necessary in cases where interactions may be present that might result in misleading conclusions. Finally, factorial designs allow the effects of a factor to be estimated at several levels of the other factors, yielding conclusions that are valid over a range of experimental conditions. Factorial design can be complicated by the introduction of blocking and counterbalancing. The latter is when people work through the conditions in different orders; it distributes order effects evenly over the conditions being compared.

There are two further important types of experimental design: between-subjects designs and within-subjects designs. We will discuss the difference between these next.

1.7.1 Between-subjects vs within-subjects designs

In classic RCTs based on between-subject design, a subject is typically allocated to only one experimental condition: either to one treatment or to one control group. This is the best approach when the conditions to be tested are few and the design very simple. With many conditions to be tested and a complex design, the solution is a between-subjects design with partial repetition of measures or a full-blown within-subjects design; in the first instance, subjects can be exposed to more than one experimental condition and in the second all subjects are used with every condition tested in the experiment, including the control one. The main advantage consists in increasing statistical power without adding more subjects but more tasks to be performed by the same individuals. Put the other way around, whereas with 120 subjects we can test a maximum of three treatments and one control in a between-subjects design, this can be multiplied by some factors (but not too much, otherwise undesired effects kick in too heavily) if each subject is used in more than one condition. Within-subjects design presents a number of challenges that are not present in between-subject design: order effects, learning effects, hypothesis guessing, and fatigue. All these effects, if not balanced, introduce noise and threaten construct validity.

Imagine a situation where all subjects go through treatment A,B,C in this exact order. Three outcomes are possible: (1) going through A before B may influence the reaction to B, which means we can no longer assume that the effect of B on the response variable X is due just to B; (2) when subjects arrive at C, they may be tired and their reactions will reflect their mental state more than the effect of treatment C; and (3) it is possible that going through A,B,C in this order enables the subjects to understand what the researcher is trying to prove. They can, thus, engage in hypothesis guessing and answer either way so to please the researchers or to refute their hypothesis. In all three cases, construct validity is weakened, and the causal effect measured may be the result of biases rather than of the treatments.

1.8 Online Experiments

The limitations of online laboratory experiments and opportunities as a field in which to conduct research has prompted researchers to develop online experiments both in academic and business settings (Veltri, 2019; Reips, 2002b; Salganik, 2019).

Giuseppe A. Veltri - 9781789906851

Of particular interest is the form of online experiments that combine the lessons learned from online surveys. The goal of the so-called 'population-based survey experiments' or PBSE, is to address this problem through research design rather than analyses, combining the best aspects of both approaches, capitalising on their strengths and eliminating many of their weaknesses.

Defined in the most rudimentary terms, a population-based survey experiment is an experiment that is administered to a representative population sample. Another common term for this approach is simply 'survey-experiment', but this abbreviated form can be misleading because it is not always clear what the term 'survey' is meant to convey. The use of survey methods does not distinguish this approach from other combinations of survey and experimental methods. After all, many experiments already involve survey methods, at least in administering pre-test and post-test questionnaires, but that is not what is meant here. Population-based survey experiments are not defined by their use of survey interview techniques—whether written or oral—nor by their location in a setting other than a laboratory. Instead, a population-based experiment *uses survey sampling methods to produce a collection of experimental subjects that is representative of the target population of interest for a particular theory, whether that population is a country, a state, an ethnic group, or some other subgroup*. The population represented by the sample should be representative of the population to which the researcher intends to extend their findings. In population-based survey experiments, experimental subjects are randomly assigned to conditions by the researcher, and treatments are administered as in any other experiment. But the participants are not generally required to show up in a laboratory in order to participate.

Table 21.2 provides a synthetic overview of the pros and cons of online experiments, where the laboratory experiment is the implicit benchmark. The main strategic advantage of an online experiment compared to a laboratory one is the increasing possibility for generalisation (external validity), greater statistical power, and quality of the data produced. Web studies, by having larger samples, usually have greater power than lab studies. Data quality can be defined by variable error, constant error, reliability, or validity. Comparisons of power and of certain measures of quality have found cases where web data are higher in quality by one or another of these definitions than are comparable lab data, though not always (Birnbaum, 2004: 825). Many web researchers are convinced that data obtained via the web can be 'better' than data obtained from students (Reips, 2000), despite the obvious advantage that the lab offers for control. The main drawback of an online experiment compared to a laboratory one is the lack of full environmental control. Participants in online experiments may answer questions and accomplish behavioural tasks in very different settings (a room with light and silence, versus one's desk at work with less light and surrounded by a lot of noise) and with different equipment (a participant may use a browser that does not correctly show a visual stimuli or may have a slow connection, thereby delaying the completion of the tasks and increasing fatigue, frustration, and 'noisy' answers). Most importantly, since lab assistants do not control participants, there are more chances that they engage in automatic answering and completion of tasks, which introduce noise in the data. This can be controlled with check questions and is less of an issue for between-subjects design with randomisation of treatments and control conditions.

There are other technical/tactical issues that can be controlled in an online experiment (multiple submissions, dropouts, self-selection—in fact almost identical to what I have already discussed about web surveys), but the main trade-off between online and lab experiments is

Giuseppe A. Veltri - 9781789906851

Table 21.2 Pros and cons of online experiments

Pros	Cons
Wider sampling access. Ease of access to a large number of demographically and culturally diverse participants	*Multiple submissions*. They can be avoided or controlled by collecting personal identification items (evidence is that multiple submission is rare).
Generalizability/quality of data. Better basis for generalising the findings to the general population and to more settings and situations (because of high external validity). Greater power and quality of data.[1]	*Less experimental control*. Variability of environments (lighting, noise, technical features of equipment used, i.e. browser type, connection speed). In absence of lab assistants' controls, participants may provide 'noisy' answers (provided without paying attention). These problems are less of an issue with between-subjects designs with random distribution of participants to experimental conditions
Better efficiency and logistics. Experimental procedures automation reduces costs and increases the uniformity of the procedure across participants.	*Self-selection*. Only interested and motivated participants may start and complete the experiment. Using the multiple site entry technique can reduce self-selection.
Detectability of motivational confounding. See opposite cell about dropout.	*Dropouts*. This is always an issue in web experiments. However, dropout can be turned into a detection device for motivational confounding.
Reduction of experimenter effects and demand characteristics.[2]	*Lack of interaction*. This may cause misunderstanding among participants but can be avoided with pre-test and pilot testing

Notes:

[1] Data quality can be defined by variable error, constant error, reliability, or validity. Comparisons of power and of certain measures of quality have found cases where web data are higher in quality by one or another of these definitions than are comparable lab data.

[2] Demand characteristic is a subtle cue that makes participants aware of what the experimenter expects to find or how participants are expected to behave. Demand characteristics can change the outcome of an experiment because participants will often alter their behaviour to conform to the experimenter's expectations.

Sources: Batinic et al. (2002); Birnbaum (2004); Dandurand et al. (2008); Gosling and Johnson (2010); Reips (2000, 2002a, 2002b); Reips and Krantz (2010).

that of exchanging greater generalizability and data power for less experimental control. It is then not surprising that experiments are often repeated with the same outcome measures both online and in laboratory settings for checking the quality and validity of data. Other advantages of online experiments are:

- the speed of data collection and experimenting around the clock;
- a high degree of automation of the experiment (low maintenance, limited experimenter effects);
- a wider sample both in size and wider geographical variations in participants.

Giuseppe A. Veltri - 9781789906851

Data collection techniques on the internet can be grouped into *server-side* and *client-side* processing. Server-side methods (a web server, often in combination with a database application), which serve web pages that can be dynamically created depending on a user's input, are less prone to platform-dependent issues, because dynamic procedures are performed on the server so that they are not subject to technical variance. Client-side methods use the processing power of the participants' computers. Therefore, time measurements do not contain error from network traffic, and problems with server availability are less likely.

In terms of disadvantages, besides the already mentioned issue of experimental control, there are other potential limitations that are apparent depending on the research topic. These include the following factors:

- web experiments' dependency on computers and networks having psychological, technical, and methodological implications;
- from a cognitive point of view, participants at computers will likely be subject to self-actualisation and other influences in computer-mediated communication;
- the temporal duration of online experiments is shorter compared to laboratory ones.

The latter point is linked to the use of online panels for recruiting participants. In my experience (see Veltri et al., 2020, Veltri and Ivchenko, 2017), the issue of obtaining generalisable samples is becoming less of a concern because of the decreasing costs of pre-recruited online panels of the general population, although worries about response rates remain open (see the section above on web surveys for further discussion on this topic). Online panels have different costs in terms of size of sample and duration of the experimental task, the most common duration being around 20 minutes. Therefore, online experiments that include tasks that will take a long time face two problems: increased costs and higher chances of reduced control over respondents. While there is no clear evidence supporting this statement, it seems reasonable to assume that online experiments exert a higher cognitive toll on respondents than laboratory settings. The reason might be that it is easier to exercise focus and attention in an environment that is designed to allow it—the laboratory—than in other environments, such as the home.

1.8.1 Online experiments as field experiments

So far, I have discussed the features of experiments having in mind RCTs as the standard, but online experiments can also take the form of field experiments (Centola, 2010; Goldstein and Hays, 2011; Hahl and Zuckerman, 2014; Ong and Wang, 2015; Parigi et al., 2017). Field experiments typically consider pools of participants from the general population acting under conditions, and making choices, in the real world. Participants are usually unaware that they are part of a social experiment until the end, thus minimising the risk of observing behaviour biased by the 'experimenter effect'. All this increases the external validity of the experimental evidence, while retaining the ability to understand the effect of a target variable that is 'manipulated' (i.e. assigned to the treatment group and not assigned to the control group). Field experiments are indeed effective in collecting evidence on actual people's behaviour. In field experiments, people are directly observed in real context of use rather than in abstract, stylised and, perhaps, far-from-real experimental tasks designed by researchers.

For example, in the commercial domain (e.g. e-commerce platform), it means that the experimenter can manipulate, one by one, the characteristics of the frame, context, and other selling conditions within which products are normally sold (e.g. colours, position, sequence,

Giuseppe A. Veltri - 9781789906851

order in which information is presented, offers, and multi-item promotions). Importantly, the simulation in the field allows researchers to assess not only the immediate reaction of consumers to minor or seemingly unperceived modifications of the selling context and frame, but also overall changes in consumers' spending behaviour in stores or online. One example of an online field experiment is titled 'Experimental evidence of massive-scale emotional contagion through social networks' (Kramer et al., 2014), in which the field was Facebook and its users (some of them) were the unaware participants in an experiment in manipulating their emotional states. This example highlights both the scientific potential of a field experiment, but also the ethical problems that can arise when the manipulation of thousands of individuals is involved.

The role of online experiments is growing largely because experiments are carried out mainly using computer-assisted tasks, and because the early obstacles to obtain representative samples is less problematic (although not completely) using pre-recruited online panels. This is true for experiments that used the online world as a 'venue' as an alternative to a laboratory. At the same time, experiments that are about native online behaviour and social phenomena have appeared. These are field experiments where the field is digital (See for example, Veltri et al., 2020). Questions about the ethical viability of these experiments are being raised with the awareness that academic researchers are not anymore alone (they probably never were) in using this methodology. Private research companies as much as big tech ones use experiments to test new features about their platform, running hundreds of field experiments per year to improve usability, prolong the interest of users, and maintain sale rates. Out of the academic ivory towers, it is very likely that all of us have been part of such an experiment when using our digital platforms of choice.

CONCLUSIONS

The use of experiments in analytically minded sociological research offers a new opportunity in light of two main changes, I believe. The first is the pervasive influence of dual model of cognition (see Chapter 7 in this book) that has come from psychological and behavioural research of the past two decades and that has led to the rise of a cognitive sociology stream of research that focuses on the study of identity, classification, memory, and in general culture. (for a dialogue between this approach and analytical sociology, see Chapter 5 in this book). This stream of sociology has been moving away from a radical opposition between emic studies of culture 'in the wild' and lab experiments. Cultural sociologists like DiMaggio (1997), Zerubavel (1997), Cerulo (2010), and Vaisey (2009) are building a bridge toward cognitive science by integrating the dual process of morality of cognition, embodiment, cognitive bandwidth, implicit bias, and emotions into cultural sociology. For example, Shepherd (2011) in particular has considered the impact of symbols and media, place, situation, and networks in implicit association test (IAT) research. This stream of sociological research requires the use of experiments because more conventional methods of sociological research are inadequate to deal with implicit cognition and attitudes, heuristics, and biases. In this case, the use of experiments is highly complementary with the generative approaches of multi-agents simulation because they can be the input of their calibration.

The increased availability of population-based survey experiments, or PBSE, opens a promising possibility. Having access to large representative samples removes one traditional obstacle of experimental research in the lab and allows exploration of more in-depth differences among

Giuseppe A. Veltri - 9781789906851

subgroups of the sample at hand. Differentiations in cognitive styles identified by experiments conducted on large and diverse samples allows avoiding extreme forms of universalism and the WEIRD bias, or the use of participants in experiments that are Western, Educated, Industrialized, Rich, and Democratic (Henrich et al., 2010). Such higher resolution segmentation can also be now carried out using abductive methods coming from computational social sciences approaches and machine learning; for example, model-based recursive partitioning (Zeileis et al., 2008).

The combination of experiments based on samples, machine learning methods, and multi-agents simulation is a likely domain of innovation that, while still in its infancy, allows analytical sociologists to dialogue with cognitive and computational sociologists (on this potential interaction, see Keuschnigg et al., 2018).

REFERENCES

Aronson, E., C. Phoebe, J. Ellsworth, M. Carlsmith, and M. Gonzales (1990), *Methods of Research in Social Psychology*, 2nd ed. New York: McGraw-Hill.

Batinic, B., U. Reips and M. Bosnjak (eds) (2002), *Online Social Sciences*, Seattle, WA: Hogrefe and Huber.

Birnbaum, M.H. (2004), "Human research and data collection via the internet", *Annual Review of Psychology*, **55**: 803–32.

Bohnet, I. (2009), "Experiments", in Peter Hedström and Peter Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press, pp. 639–65.

Bronfenbrenner, U. (1977), "Toward an experimental ecology of human development", *American Psychologist*, **32**: 513–31.

Brunswik, E. (1943), "Organismic achievement and environmental probability", *Psychological Review*, **50**: 255–72.

Campbell, D.T. (1957), "Factors relevant to the validity of experiments in social settings", *Psychological Bulletin*, **54** (4): 297–312.

Campbell, D.T. and Stanley, J. (1966), *Experimental and Quasi-Experimental Evaluations in Social Research*, Chicago, IL: Rand McNally.

Centola, D. (2010), "The spread of behavior in an online social network experiment", *Science*, **329** (5996): 1194–7.

Cerulo, K.A. (2000), "Culture and cognition: New approaches to traditional concepts", *Poetics*, **28** (1), 1–90.

Cohen, B.P. (1989), *Developing Sociological Knowledge: Theory and Method*, Chicago: Nelson–Hall.

Cohen, B.P. (1997), "Beyond Experimental Inference: A Decent Burial for J.S. Mill and R.A. Fisher", in J. Szmatka, J. Skvoretz and J. Berger (eds), *Status Network and Structure: Theory Development in Group Processes*, pp. 71–86, Stanford, CA: Stanford University Press.

Cook, T. and D. Campbell (1979), *Quasi-Experimentation: Design and Analysis for Field Settings*, Chicago, IL: Rand McNally.

Cronbach, L.J. (1982), *Designing Evaluations of Educational and Social Programs*, San Francisco, CA: Jossey-Bass.

Dandurand, F., T. Shultz and K. Onishi (2008), "Comparing online and lab methods in a problem-solving experiment", *Behavior Research Methods*, **40** (2): 428–34.

Deaton, A. (2009), "Instruments of development: Randomisation in the tropics, and the search for the elusive keys to economic development", The Keynes Lecture, British Academy, London.

DiMaggio, P. (1997), "Culture and cognition", *Annual Review of Sociology*, **23**, 263–87.

Epstein, J.M. (1999), "Agent-based computational models and generative social science", *Complexity*, **4** (5), 41–60.

Folkvord, Frans, Cristiano Codagnone, Francesco Bogliacino and Giuseppe A. Veltri (2019), "Experimental evidence on measures to protect consumers of online gambling services", *Journal of Behavioral Economics for Policy*, **3** (1): 20–9.

Goldstein, N.J. and N.A. Hays (2011), "Illusory power transference: The vicarious experience of power", *Administrative Science Quarterly*, **56** (4): 593–621.

Gosling, S. and J. Johnson (eds) (2010), *Advanced Internet Methods in the Behavioral Sciences*, Washington, DC: American Psychological Association.

Hahl, O. and E.W. Zuckerman (2014), "The denigration of heroes? How the status attainment process shapes attributions of considerateness and authenticity", *American Journal of Sociology*, **120** (2): 504–54.

Harrison, G. and J. List (2004), "Field experiments", *Journal of Economic Literature*, **42**: 1013–59.

Henrich, J., S.J. Heine et al. (2010). "The weirdest people in the world", *Behavioral and Brain Sciences*, **33**: 2–3.

Howitt, D. and D. Cramer (2011), *Introduction to Research Methods in Psychology*, Harlow, UK: Pearson Education.

Giuseppe A. Veltri - 9781789906851

Keuschnigg, Marc, Niclas Lovsjö and Peter Hendström (2018), "Analytical sociology and computational social science", *Journal of Computational Social Science*, **1**: 3–14
Kiesler, C.A., B.E. Collins and N. Miller (1969), *Attitude Change*, New York: Wiley.
Kramer, A.D.I., J.E. Guillory and J.T. Hancock (2014), "Experimental evidence of massive-scale emotional contagion through social networks", *Proceedings of the National Academy of Sciences*, **111** (24): 8788–90.
Manzo, G. (2014a), "Potentialities and limitations of agent-based simulations: An introduction", *Revue française de sociologie*, **55** (4): 653–88. https://doi.org/10.3917/rfs.554.0653
Manzo, G. (ed.) (2014b), *Analytical Sociology: Actions and Networks*, Chichester, UK: Wiley-Blackwell.
Martin, J.L. (2011), *The Explanation of Social Action*, Oxford: Oxford University Press.
McGraw, K. and V. Hoekstra (2010), "Experimentation in Political Science", in J.M. Box-Steffensmeier, H.E. Brady and D. Collier, *The Oxford Handbook of Political Methodology*, Oxford: Oxford University Press.
Montgomery, D.C. (2009), *Introduction to Statistical Quality Control*, Wiley.
Neisser, U. (1976), *Cognition and Reality: Principles and Implications of Cognitive Psychology*, New York: WH Freeman and Company.
Neyman, J. (1923), "On the application of probability theory to agricultural experiments: Essay on principles", *Statistical Science*, **5** (4): 465–80.
Ong, D. and J. Wang (2015), "Income attraction: An online dating field experiment", *Journal of Economic Behavior and Organization*, **111**: 13–22.
Parigi, P., J.J. Santana and K.S. Cook (2017), "Online field experiments: Studying social interactions in context", *Social Psychology Quarterly*, **80** (1): 1–19.
Reips, U.D. (2000), "The Web Experiment Method: Advantages, Disadvantages, and Solutions", in M. Birnbaum (ed.), *Psychological Experiments on the Internet*, San Diego, CA: Academic Press.
Reips, U.D. (2002a), "Theory and Techniques of Conducting Web Experiments", in B. Batinic, U. Reips and M. Bosnjak (eds), *Online Social Sciences*, Seattle, WA: Hogrefe and Huber.
Reips, U.D. (2002b), "Standards for internet-based experimenting", *Journal of Experimental Psychology*, **49** (4): 243–56.
Reips, U.D. and J. Krantz (2010), "Conducting True Experiments on the Web", in S. Gosling and J. Johnson (eds), *Advanced Internet Methods in the Behavioral Sciences*, Washington DC: American Psychological Association.
Rubin, D.B. (1974), "Estimating causal effects of treatments in randomised and non-randomised studies", *Journal of Educational Psychology*, **66** (5): 688–701.
Salganik, M. (2019), *Bit by Bit: Social Research in the Digital Age*, Princeton: Princeton University Press.
Scheidt, R. (1981), "Ecologically-valid inquiry: Fait accompli?", *Human Development*, **24**: 225–8.
Shadish, W., T. Cook and D. Campbell (2002), *Experimental and Quasi Experimental Designs for Generalised Causal Inference*, Boston, MA: Houghton Mifflin.
Shepherd, H. (2011), "The cultural context of cognition: What the implicit association test tells us about how culture works", *Sociol. Forum*, **26**, 121–43.
Simon, H.A. (1954), "Spurious correlation: A causal interpretation", *Journal of the American Statistical Association*, **49**: 467–92.
Sobel, M.E. (1995), "Causal Inference in the Social and Behavioral Sciences", in G. Arminger, C.C. Clogg and M.E. Sobel (eds), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, Boston, MA: Springer.
Solomon, P.L., M.M. Cavanaugh and J. Draine (2009). *Randomized Controlled Trials: Design and Implementation for Community-Based Psychosocial Interventions*, New York: Oxford University Press.
Vaisey, S. (2009), "Motivation and justification: A dual-process model of culture in action", *American Journal of Sociology*, **114**: 1675–715.
Veltri, G.A. (2019), *Digital Social Research*, Cambridge: Polity Press.
Veltri, G.A. and A. Ivchenko (2017), "The impact of different forms of cognitive scarcity on online privacy disclosure", *Computers in Human Behavior*, **73**, 238–46. https://doi.org/10.1016/j.chb.2017.03.018
Veltri, G.A., F. Lupiáñez-Villanueva, F. Folkvord, A. Theben and G. Gaskell (2020), "The impact of online platform transparency of information on consumers' choices", *Behavioural Public Policy*, 1–28. https://doi.org/10.1017/bpp.2020.11
Woodward, J. (2000), "Explanation and invariance in the special sciences", *British Journal for the Philosophy of Science*, **51**: 221–2.
Yule, G.U. (1896), "On the correlation of total pauperism with proportion of out-relief. II: Males over 65", *Economic Journal*, **6**: 613–23.
Zeileis, A., T. Hothorn and K. Hornik (2008), "Model-based recursive partitioning", *Journal of Computational and Graphical Statistics*, **17** (2): 492–514.
Zerubavel, E. (1997), *Social Mindscapes: An Invitation to Cognitive Sociology*, Cambridge, MA: Harvard University Press.

Giuseppe A. Veltri - 9781789906851

22. Game-theoretic models[1]

Wojtek Przepiorka

1. INTRODUCTION

This chapter focuses on game-theoretic models of social dilemmas. A social dilemma is, as the name suggests, a situation in which several agents face a decision conflict because of their interdependence. Social dilemmas are thus to be distinguished from ethical dilemmas. The latter comprise decision-making situations in which a single agent is in conflict with themselves because none of the available decision options is compatible with their ethical values. Although there is a clear connection between social and ethical dilemmas (see, e.g. Green 2013), this chapter focuses on social dilemmas.

There are numerous definitions of social dilemmas (Dawes 1980; Diekmann and Przepiorka 2016; Kollock 1998; Raub et al. 2015; Van Lange et al. 2013). This chapter uses the definition from the German-language Wikipedia entry on the term (Wikipedia Contributors 2018; translated by the author): Social dilemmas exist if the best possible pursuit of the involved agents' individual interests leads these agents into a state that puts them in a worse position than solutions in which the best action to maximize individual interests is abandoned. This definition expresses the conflict between individual and collective rationality inherent in social dilemmas (Rapoport 1974), without assuming particular interaction situations and theoretical solution concepts.

This chapter discusses social dilemmas from a game-theoretic perspective, because this allows one to describe precisely how the best possible pursuit of individual interests puts the agents involved in a worse position than other decision alternatives. The focus will be on the application of game theory rather than the mathematical derivation of game-theoretic models. Moderately technical introductions to game theory are offered by, for example, Camerer (2003), Osborne (2009) or Rasmusen (2007). An example-rich, but technically more involved introduction to game theory, is offered by Gintis (2009). A game-theoretic discussion of social dilemmas can also be found in Raub et al. (2015).

In the three sections following this introduction, social dilemmas that occur in three areas of application are discussed: Cooperation, coordination and conflict. The game-theoretic modelling of social dilemmas allows one to describe interaction situations that comprise cooperation problems, coordination problems and conflicts of interest. Since such problems of social interaction occur in numerous situations, the game-theoretic modelling of social dilemmas allows one to describe and explain real and sociologically relevant phenomena. The last two sections before the Summary section address two topics that further extend the applicability of game-theoretic modelling of social dilemmas for sociological theory building: Repetition and Information. Repetition accounts for the fact that frequently agents encounter each other repeatedly in the same interaction situations, which changes their incentives to choose certain

[1] This chapter is a translated and substantially extended version of Przepiorka (2020).

Wojtek Przepiorka - 9781789906851

decision options as compared to one time–only encounters. The section on Information relaxes the assumption that the agents involved in an interaction are always perfectly informed about each other's preferences and constraints. In fact, many social dilemmas arise because of so-called information asymmetries among agents.

Game theory is a tool to model and discover social mechanisms. However, game theory makes no pretence to be the only such tool. It should be considered as one (important) tool among many in the analytical sociologist's tool box (see, e.g. Hedström and Bearman 2009; Manzo 2014, and the other chapters in this book). The aim of this chapter is to show how the tool can be used by engaging the reader's interest in the game-theoretic modelling of social interactions in general and social dilemmas in particular. The chapter therefore uses examples to connect the formal game-theoretic models with everyday life occurrences of social dilemmas. However, the attentive reader will notice that there are conceptual gaps between the models and the examples used for illustration. This does not mean, however, that game theory is unsuitable for precisely describing real social dilemmas. The sections on Repetition and Information provide insight into how some important gaps between model and reality could be reduced. Still, a comprehensive game-theoretic treatment of social dilemmas is not possible in this chapter. It is therefore recommended that the interested reader consider the further literature cited in this chapter.

2. COOPERATION

The best-known example of a cooperation dilemma is the Prisoner's Dilemma (PD). In Figure 22.1a the two-agent PD is shown in so-called normal form. In the PD, both agents decide simultaneously and independently of each other whether to cooperate (C) or to defect (D). Each of the four decision combinations is assigned a payoff pair, where the value before the comma indicates the payoff of agent 1 and the value after the comma indicates the payoff of agent 2. The PD is defined by the ranking of these payoffs. A PD exists if the following applies: $T > R > P > S$. It is therefore always better to choose D; if the other agent chooses D, one gets P (which is more than S), and if the other chooses C, one gets T (which is more than R). If both agents follow this logic and choose D, they get P each. Mutual defection is also the Nash equilibrium in the PD; a Nash equilibrium exists if no agent has an incentive to deviate unilaterally from their chosen strategy.

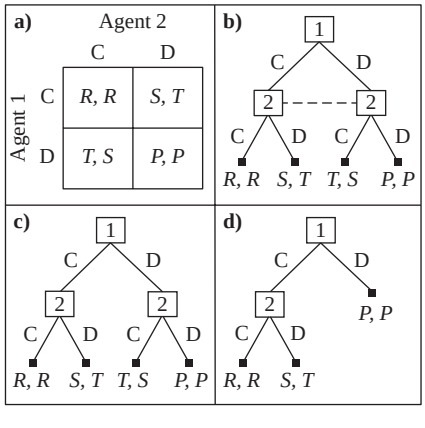


Figure 22.1 Prisoner's dilemma and trust dilemma

Wojtek Przepiorka - 9781789906851

The PD clearly expresses the characteristics of a social dilemma. The best possible pursuit of individual interests (i.e. the choice of D by both agents) leads both agents into a state (both receive *P*) in which they are worse off than in the state they would have reached by choosing C (both receive $R > P$). In cooperation dilemmas, it can be said that the Pareto-efficient state, in which none of the agents can be made better off without making another agent worse off, does not correspond to the state in the Nash equilibrium.

The PD is often used to describe so-called commons problems (Ostrom 1990) or the free-rider problem occurring with public goods. A commons problem concerns the use of a scarce resource from which nobody can be excluded. However, if all agents (usually more than two) use the resource at their discretion, the resource will be exhausted or the costs of its use will exceed the benefits. Overfishing of the world's oceans, congested roads in urban areas, poor air quality or excessive noise levels in the quiet compartment of a train are examples of commons problems. Although it is possible to solve commons problems by means of fishing quotas, driving bans and appeals to behavioral norms, such approaches often create second-order social dilemmas (Heckathorn 1989; Voss 2001; Yamagishi 1986). This means that in order to solve the first-order cooperation dilemma, a second-order social dilemma must first be overcome.

In a functioning constitutional state, driving bans can be enforced by the police, i.e. a central authority, and the taxpayer bears the costs. However, enforcing compliance with fishing quotas requires the cooperation of at least two states or regions, which are not necessarily linked by common legislation. The emergence of supranational institutions such as the European Union or the United Nations has contributed significantly to the resolution of second-order social dilemmas at international level. However, the multilateralism that has come under pressure in recent years (e.g. the US's withdrawal from the Paris Climate Agreement, the UK's withdrawal from the EU) is reducing hopes for a fair solution to pressing global problems such as climate change or tax avoidance by international corporations.

The examples just mentioned show that cooperation dilemmas (and other social dilemmas) are not limited to 2-agent interactions between individuals. A cooperation dilemma can be difficult to overcome only in large groups, because the large number of agents makes communication and agreement on a solution difficult. PD-like social dilemmas are nevertheless often encountered in dyads, i.e. in interactions between two agents. Probably the most prominent example is the trust dilemma (TD) (Dasgupta 1988; Kreps 1990), which can be easily derived from the PD (Diekmann et al. 2009).

Figure 22.1b shows the PD in extensive form, i.e. as decision tree. The game logic remains the same. Both agents decide simultaneously and independently of each other whether they choose C or D. The fact that both agents do not know what the other agent chooses is indicated by the dashed line. The dashed line connects the decision nodes of agent 2 and means that agent 2 does not know how agent 1 has decided (agent 2 does not know which of the two decision nodes they are at) when they make their decision. If the dashed line is omitted as in Figure 22.1c, the simultaneous PD becomes a sequential PD. In the sequential PD agent 1 chooses first and agent 2 learns how agent 1 has decided before agent 2 makes their choice. The sequential PD still has the characteristics of a social dilemma. This can easily be seen by analyzing the game from below, i.e. by backward induction. Agent 2 will always choose D because D will bring them the higher payoff, no matter what agent 1 does. Since agent 1 knows the game as well as the reader does, agent 1 will anticipate the choice of agent 2 and in turn choose D. This is because even for agent 1, it is best to choose D when it is clear that agent 2 will choose D. From the sequential PD the TD can now be derived by eliminating the

Wojtek Przepiorka - 9781789906851

option of agent 2 in case agent 1 chooses D (Figure 22.1d). In the TD, if agent 1 chooses D, the interaction is finished and both agents receive *P*.

The TD is often used to model the trust problem inherent in sequential exchange (Güth and Ockenfels 2003). Agent 1 and agent 2 possess something that the respective other appreciates more (e.g. money and goods). Through an exchange both would be better off since they would receive what they desire more at that moment. However, the exchange is sequential and agent 1 has to pay agent 2 in advance (e.g. transfer the money). Then agent 2 can decide whether to reward the advance payment of agent 1 or not. That is, agent 2 decides whether to reciprocate (hand over the goods) or not (keep the goods and the money).

As with the sequential PD, the Nash equilibrium can be derived by backward induction. Agent 2 has an incentive to keep the money received from agent 1 without giving them the promised goods ($T > R$). Since agent 1 anticipates the choice of agent 2, they in turn will choose D ($P > S$). The exchange does not take place, so both agents lose the gains from trade ($R - P$). But how can it be that tens of thousands of people trade with each other every day? Especially on anonymous online markets, the occurrence of barter transactions is in need of explanation. Basically, there are three mechanisms that help to solve trust dilemmas in markets: social, legal and organizational (Diekmann and Przepiorka 2019).

Social and economic exchange is often embedded in the networks of relationships of the agents involved in the exchange (Granovetter 1992). Due to different mechanisms, the social embeddedness of the agents can mitigate the trust dilemma inherent in sequential exchange (Buskens and Raub 2013). For example, the social embeddedness of the agents can result in a high probability that they will meet again in a similar exchange situation. In this case, the agents (and especially the trusted party) have an incentive to behave cooperatively in order not to lose the gains from future exchange interactions. If the so-called shadow of the future (Axelrod 1984) is large, and the probability that, after a bad experience, the trusting agent will trust the trusted agent in the future is small, then it is in the trusted agent's own interest to behave cooperatively and reward the trust given to him (Hardin 2002). The Repetition section in this chapter expands on this mechanism.

The social embeddedness of the agents also mitigates the TD in that third parties learn about the outcome of an interaction through their own observation or word-of-mouth communication and may in future refuse to trust an uncooperative trusted agent as well as those agents close to them (Sosis 2005). The latent threat that uncooperative behavior may result in a negative sanction creates an incentive for cooperative behavior (Diekmann and Przepiorka 2015; Fehr and Gächter 2002). However, the "community" solution to the TD reaches its limits when a social and economic exchange takes place over long geographical distances and between agents who do not meet again. Within national borders and a corresponding jurisdiction, social and economic exchange can be promoted by enforceable contracts that mitigate the trust dilemma. The embeddedness of the agents in formal institutions shifts the trust dilemma partly out of the dyad into the relationship between citizen and state. In other words, agents have to have less trust in each other in order to enter mutually beneficial exchanges as long as they can trust state institutions to enforce their rights, if necessary. However, the TD between agents can never be completely resolved by state institutions; contracts cannot cover every eventuality, and legal action in court is often associated with high costs.

If the social and institutional embeddedness of the agents interested in an exchange is not given, organizational mechanisms can contribute to mitigating the TD. The potential of organizational mechanisms has increased greatly with the spread of the internet over the last 25

Wojtek Przepiorka - 9781789906851

years. Electronic reputation systems, in particular, have contributed to making cross-border exchanges between anonymous agents possible on an unprecedented scale (Diekmann and Przepiorka 2019). The reputation mechanism is based on the fact that agents must behave cooperatively in order to build and maintain a good reputation. A good reputation promises higher exchange gains which, however, will only accrue in the future. Agents who are only interested in the short-term exploitation of their exchange partners will, therefore, shy away from the costs associated with building a good reputation (Przepiorka 2013; Shapiro 1983). The success of illegal online markets in the "Dark Net" is an example of how organizational mechanisms such as electronic reputation systems can solve cooperation dilemmas, even without the social and institutional embeddedness of the interacting agents (Przepiorka et al. 2017).

Removing the incentives for mutual defection through social, legal and organizational mechanisms is not always sufficient to solve a social dilemma. Coordination problems often remain, which in turn can constitute social dilemmas. For example, two countries can conclude an agreement that removes trade barriers and helps both countries achieve greater prosperity. But there are countless ways to define such an agreement, and what is considered cooperative in the eyes of politicians in one country can be perceived as a deficiency by politicians in the other. The two countries must therefore first solve a coordination dilemma and agree on what is considered cooperative (McAdams 2009; see also Garrett and Weingast 1993). As the next section will show, many areas of life are characterized by coordination dilemmas.

3. COORDINATION

In contrast to cooperation dilemmas, coordination dilemmas are dilemmas in which the individual interests of the agents are fully or partially aligned (the distinction between fully and partially aligned interests is discussed in more detail below). The social dilemma is now that there is more than one way to pursue one's individual interests in the best possible way. Since the assumption of there being no communication between agents still holds, these agents may fail to consciously coordinate their actions, which sometimes leads them into a worse state. This can be illustrated by the coordination dilemma in Figure 22.2a.

Figure 22.2a shows a pure coordination game in which two agents decide on an action a or b. If both choose the same action (both choose a, or both choose b), they receive the same positive payoff. If both agents decide on different actions (agent 1 chooses a and agent 2 chooses b, or vice versa), both go away empty-handed. Unlike a cooperation dilemma such as the PD, here the states in which both agents maximize their individual payoffs are also Nash equilibria. Moreover, in this case the individual interests of the two agents are completely aligned. So it is not the case that one agent would prefer one Nash equilibrium and the other agent would prefer the other equilibrium. However, since the two agents make their decisions simultaneously and independently of each other, there are no clues that could help them coordinate their actions in order to arrive at one of the Nash equilibria.

A clue would arise if, for both agents, one of the two Nash equilibria would yield a higher payoff (Schelling 1960). Figure 22.2b shows such a coordination dilemma if $c = 0$. This coordination dilemma also has two Nash equilibria, but one of them is Pareto efficient because it yields a higher payoff for both agents. In this case, the social dilemma is significantly mitigated but not eliminated; there is still a second Nash equilibrium in which both agents choose b.

Wojtek Przepiorka - 9781789906851

a)

Agent 1 \ Agent 2	a	b
a	1, 1	0, 0
b	0, 0	1, 1

b)

Agent 1 \ Agent 2	a	b
a	2, 2	0, *c*
b	*c*, 0	1, 1

c)

Agent 1 \ Agent 2 \ Agent 3	a: a	a: b	b: a	b: b
a	2, 2, 2	0, 1, 0	0, 0, 1	0, 1, 1
b	1, 0, 0	1, 1, 0	1, 0, 1	1, 1, 1

Figure 22.2 Coordination dilemmas

A small change in the payoff matrix illustrates why the Pareto-inefficient Nash equilibrium cannot be ignored.

With $c = 1$, the attraction of the Pareto-inefficient Nash equilibrium increases, because now the choice of b, regardless of what others do, gives a safe payoff of 1, while the choice of a can give a payoff of 2 or 0. However, this aspect requires that one assumes that agents sometimes make mistakes when choosing their preferred action (Selten 1975), i.e. they choose b even though they would have wanted to choose a, and/or some agents are risk-averse and want to play it safe, and therefore choose b.

Many situations described as PD (i.e. as interactions between two prisoners who decide independently whether to keep quiet or reveal) would actually be more aptly formulated in terms of the coordination dilemma in Figure 22.2b with $c = 1$ (McAdams 2009). This coordination dilemma is also called the assurance dilemma (AD), because both agents would like to have an assurance from the other that a will be chosen, but there can be no certainty that such a promise will be kept. This aspect is explained with the following example and, at the same time, it is shown how a social dilemma in which three agents are involved can be modelled.

Three bosses of competing companies have made illegal price agreements that undermine the consumer surplus and are accused of this by the public prosecutor. However, the public prosecutor has no evidence of the existence of a cartel and is, therefore, dependent on at least one of the bosses to spill the beans. The three bosses are summoned and find themselves in the situation formalized in Figure 22.2c. Note that Figure 22.2c does not show two different interactions, but the payoff matrix of the 3-agent game. Since three players each have the same two decision options (keep quiet or reveal), this results in $2^3 = 8$ decision combinations, each of which is assigned a payoff triplet. Unlike the 2-agent case, each cell of the payoff matrix contains three numbers. The first number is the payoff of agent 1, the second number is the payoff of agent 2 and the third number is the payoff of agent 3.

All three decide at the same time and independently of each other whether they choose a (keep quiet) or b (reveal). If an agent reveals (chooses b), they receive a payoff of 1 (for example, a lower penalty), regardless of what the others decide. If an agent remains silent (chooses a), their payoff depends on what the other two do. If at least one of the other agents

Wojtek Przepiorka - 9781789906851

reveals, agent 1 goes away empty-handed (agent 1 gets the maximum penalty). If both remain silent, the payoff of agent 1 (as well as the payoff of the other players) is 2 (the cartel remains unpunished). This is a coordination dilemma because revealing is not a dominant strategy (so it is not always better to reveal) and there are several (two) Nash equilibria. The dilemma exists because even if the three company bosses agree in advance to keep quiet, there is no guarantee that one of them will not get cold feet and reveal all anyhow, maybe even out of fear of getting the maximum penalty if one of the other bosses reveals.

The AD is a collective good dilemma that requires the "inaction" of all the agents involved in order to produce the collective good (i.e. a benefit for all group members). In the AD, a single deviator (who chooses b) can destroy the collective good, but suffers a smaller loss than the other agents. In the so-called volunteer's dilemma (VOD), in contrast, only one volunteer is needed to produce the collective good for all agents involved (Diekmann 1985). The VOD, which is discussed next, is, so to speak, point-symmetrical to the AD.

The VOD is a collective good game where the effort of only one agent is required to produce the collective good for all the agents involved. The VOD occurs in many situations in life: Who does the dishes after dinner in the shared flat? Who gets up to calm the crying infant? Who intervenes when a resident in the apartment block disturbs the night's rest? Who takes over the management of a new project at the workplace? Who abstains from drinking alcohol to drive everyone home safely after a party? Figure 22.3a shows the 3-agent VOD in normal form.

In the VOD shown in Figure 22.3a, three agents decide simultaneously and independently of each other whether they will make an effort for the group and produce the collective good (choose v) or not (choose ¬v). For example, if agent 1 chooses v, they receive a payoff of 1, regardless of what the other two agents do. If agent 1 chooses ¬v, then their payoff depends on what the others do. If at least one of the other agents chooses v, then agent 1 will receive a payoff of 2, but if all agents choose ¬v, all will go away empty-handed, since in this case the collective good will not be produced. Finally, note that, unlike in other collective goods games, the contribution of more than one agent does not increase the payoff of the other agents. In the VOD, it is therefore unnecessary for more than one agent to choose v.

a)

		Agent 3: v		Agent 3: ¬v	
		Agent 2: v	Agent 2: ¬v	Agent 2: v	Agent 2: ¬v
Agent 1	v	1, 1, 1	1, 2, 1	1, 1, 2	1, 2, 2
Agent 1	¬v	2, 1, 1	2, 2, 1	2, 1, 2	0, 0, 0

b)

		Number of other agents that choose v: 0	1	...	$n - 1$
Agent i	v	$U - K$	$U - K$	$U - K$	$U - K$
Agent i	¬v	0	U	U	U

Figure 22.3 Volunteer's dilemma

Wojtek Przepiorka - 9781789906851

More generally (and perhaps even more intuitively), the *n*-agent VOD can be represented as in Figure 22.3b. Since it is assumed that the VOD is symmetric, i.e. the position of an agent does not determine their payoff, the VOD can be viewed from the perspective of any agent *i*. As in the 3-agent case, an agent *i* must decide between v and ¬v. If agent *i* chooses v, they incur costs *K* for the production of the collective good, which results in a payoff of *U* for each agent involved. In this case, each agent receives the payoff *U*, regardless of whether they chose v or ¬v. But if agent *i* chooses ¬v, then their payoff depends on how many other agents choose v. If none of the other agents chooses v, the collective good is not produced and all go away empty-handed. If at least one of the other agents chooses v, agent *i* receives payoff *U* without having to pay *K* for the production of the collective good.

Unlike the AD, where the number of Nash equilibria is two, independent of the number of agents, the number of Nash equilibria in the VOD corresponds to the number of agents involved. In a Nash equilibrium in the VOD, exactly one agent chooses v and the other $n - 1$ agents choose ¬v. The VOD is a hybrid of a cooperation and a coordination dilemma. It is a cooperation dilemma because one agent has to bear a cost to produce the collective good. But, unlike the PD, the benefits of the collective good exceed the costs of its production ($U - K > 0$). It is a coordination dilemma because the agents have no clue that would allow them to tacitly agree on one of the *n* Pareto-optimal Nash equilibria. Who should be the one to take on the task of producing the collective good?

In a situation in which a group of agents is confronted with a VOD (e.g. the residents of an apartment block who are disturbed by the loud music of a resident, and with a call to the police could re-establish the collective good of "silence"), there can be a diffusion of responsibility (Darley and Latané 1968). If all agents expect one of the others to perform the task (e.g. to make the call to the police) and produce the collective good, the collective good is only produced after a costly delay or not at all.

As mentioned above, coordination dilemmas can be mitigated by making it easier for agents to select one of the existing Nash equilibria (i.e. tacit agreement) through asymmetric payoffs (Harsanyi and Selten 1988). In order to break up existing cartels and prevent the formation of new ones, some countries grant immunity from prosecution or a substantial reduction in penalties to cartel members who are the first to give up the cartel (Bigoni et al. 2015). The diffusion of responsibility in the VOD can be avoided by attributing lower costs or higher benefits to one agent for the creation of the collective good (Olson 1965; Przepiorka and Diekmann 2018). In an asymmetric VOD with a "strong" agent, the Nash equilibrium in which the strong agent creates the collective good alone is also the Pareto-optimal Nash equilibrium (Diekmann 1993).

In the vast majority of cases, social dilemmas are characterized by asymmetries in the payoffs of the agents. In an apartment block, the next-door neighbor of the resident who disturbs the night's rest has, under otherwise constant conditions, both the lowest cost and the greatest benefit of calling the police or intervening themselves (Przepiorka and Berger 2016). However, symmetrical coordination dilemmas also find a solution when they are encountered repeatedly by the same agents. In repeated interactions, the possibility arises of taking turns in the production of the collective good (Lau and Mui 2012; Diekmann and Przepiorka 2016). In the group of friends that goes to town every weekend to celebrate, the drivers can take turns. However, it becomes difficult if only one of them has a driving license (so the social dilemma is very asymmetric). If the group tacitly agrees each time that the member with the lowest cost or the greatest benefit will create the collective good, a conflict can arise because over time the payoffs of the agents will be unequally distributed.

Wojtek Przepiorka - 9781789906851

Coordination dilemmas, in which the individual interests of the agents are only partially aligned, are well suited to modelling conflicts. The next section discusses social dilemmas that arise when multiple Nash equilibria lead to asymmetries in agents' payoffs, and distributional conflicts can, therefore, arise.

4. CONFLICT

The game shown in Figure 22.4a is called "battle of the sexes." It expresses the conflict between two agents who want to spend time together but disagree on how to spend time together. This is illustrated by the following example. Agent 1 likes cinema more than theatre, while agent 2 has the opposite preference order. But regardless of their preferences, both like to spend time together rather than alone. Both agents now decide simultaneously and independently of each other whether they choose h (cinema) or z (theatre). The game has two Pareto-optimal Nash equilibria; one in which the agents go to the cinema together (choose h) and one in which they go to the theatre together (choose z). But if both follow their individual preferences and choose cinema (agent 1) or theatre (agent 2), both will watch the respective performances without the other, which is the least enjoyable for both. In order for the evening not to be a failure, one of the agents must refrain from maximizing their individual interests. As with the VOD, the question of who should make the "sacrifice" arises, because if both practice self-denial and choose what the other prefers, the evening will not succeed either. The willingness to renounce can, in combination with communication, lead to a solution of the dilemma. But if neither of them is willing to give in voluntarily, a coin toss, the outcome of which both agree to respect, can help solve the dilemma. The dilemma can be solved more easily if the same two agents meet more often. In repeated encounters, both can take turns in renouncing and go to the cinema first and the theatre next time. This solution has the consequence that, over time, both of them achieve the same benefits on average.

Many conflict situations are characterized by an even greater discrepancy between the individual interests of the agents involved than in the game shown in Figure 22.4a. Figure

a)

Agent 1 \ Agent 2	h	z
h	2, 1	0, 0
z	0, 0	1, 2

b)

Agent 1 \ Agent 2	h	z
h	–.5, –.5	4, 0
z	0, 4	2, 2

c)

Agent 1 \ Agent 2	h	z
h	$.5w - c$ $.5w - c$	w 0
z	0 w	$.5w$ $.5w$

d)

Agent 1 \ Agent 2	h	z
h	$w - c$ $-c$	w 0
z	0 w	$.5w$ $.5w$

Figure 22.4 Conflict dilemmas

Wojtek Przepiorka - 9781789906851

22.4b shows the so-called chicken dilemma (CD) in normal form (Rapoport and Chammah 1966). In the CD, two agents decide simultaneously and independently of each other whether to choose h or z. If both choose h, both lose, which is expressed by the negative payoff of -0.5. If one agent chooses h and the other z, the former gets 4 and the latter gets nothing. Finally, if both chose z, both receive 2. Although the state that occurs when both choose z is Pareto optimal (no agent can be made better off without making another agent worse off), this state is not a Nash equilibrium. Both agents individually have an incentive to deviate from their strategy and choose h. The CD thus has two Nash equilibria; one in which agent 1 chooses h and agent 2 chooses z, and one in which agent 1 chooses z and agent 2 chooses h. But if both agents follow their individual incentives and choose h, they find themselves in the worst possible state. This is also the difference to the PD. In the PD, the worst possible state occurs only for the agent who cooperates while the other one defects. In the CD (and in the VOD) all-sided "defection" leads to the worst possible state.

The chicken dilemma is so-called because it is often illustrated with the example of two rivals racing toward each other in their cars. If neither of them swerves in time (both choose h), they crash and lose. If one of them swerves (chooses z) and the other continues (chooses h), the former is the "chicken" and the latter the winner of this competition. If both swerve, there is no winner, but there is no chicken either, and no dead or seriously injured. The CD was also used to analyze the Cuban Missile Crisis (Poundstone 1992). In 1962, the US and the Soviet Union were on the verge of a nuclear war with the Cuban Missile Crisis, which did not happen, but only because the then Kremlin chief, Khrushchev, gave in and ordered the ships with medium-range missiles sent to Cuba to turn back. However, the interaction was sequential, because the American president, Kennedy, had previously activated American medium-range missiles based in Turkey, which made the Soviets feel threatened.

Conflicts often involve the division of a scarce but desired resource. This can be a packet of cigarettes, the attention of a person, market share or territories rich in natural resources. If two agents are in dispute over a resource, their interaction can be described in the game shown in Figure 22.4c. In this game, both agents decide whether to enter into a costly "fight" for the resource (choose h) or not (choose z). If both choose h, a confrontation occurs, in which with a probability of 0.5 one wins the resource (w), the other goes away empty-handed and both have to bear the costs of the confrontation (c). If one agent chooses confrontation (h) while the other evades it (chooses z), the former wins the entire resource without a fight, the latter goes away empty-handed and neither of them suffers losses that a fight would have caused. Finally, there is the possibility that both of them agree to divide the resource peacefully fifty-fifty (choose z) and thus save the costs of a fight.

If $w = 4$ and $c = 2.5$, one gets the same payoff matrix as in Figure 22.4b, so the game is a CD. This also means that the peaceful sharing of the resource is Pareto optimal but not a Nash equilibrium. Since both players do not know how a fight would end (both have a 50% chance of winning the fight), both have an incentive to choose h. If both now follow their individual incentives, one of the agents will make a profit ($w - c$) and the weaker one will suffer a loss ($-c$), but in expectation ($0.5[w - c] - 0.5c = 0.5w - c$) the mutual choice of h still leads to the worst possible condition for both agents ($0.5w - c < 0$). The fact that one of the two agents emerges as the winner of the conflict with a probability of 0.5 if both choose h is derived from the following assumptions: (i) one of the agents is stronger than the other, (ii) agents do not know who the stronger agent is, but know that either of them is the stronger with equal probability, and (iii) strength determines the outcome of the

Wojtek Przepiorka - 9781789906851

conflict. In this sense, the game shown in Figure 22.4c is one with incomplete information (see, e.g. Rasmusen 2007).

So far, it has been assumed that all agents know about the preference order of all agents involved. In the case of the game shown in Figure 22.4c, however, it is implicitly assumed that the agents do not know how strong they are and how strong their opponent is, but that there are differences in the strength of the agents that determine the outcome of the interaction. If the interaction were one with complete information, it would correspond to the game shown in Figure 22.4d. In this game agent 1 is the stronger agent, agent 2 the weaker agent, and there is common knowledge about this. That agent 1 is the stronger agent can be seen from the fact that agent 1 would get $w - c$ and agent 2 would get $-c$ if both chose h. This game has a Pareto-efficient Nash equilibrium in which agent 1 chooses h and agent 2 chooses z. In other words, in the Nash equilibrium, the stronger agent seeks confrontation, the weaker one evades it, the costly confrontation does not happen, and the interaction ends with an unequal distribution of the desired resource (the strong agent gets everything, the weak agent gets nothing).

An important insight from this analysis is that resource-consuming conflicts can arise, among other things, because agents are not aware of their own strengths and the strengths of their opponents. However, the conflicts resulting from this lack of information then bring the necessary information to light. Hierarchies emerge that give the stronger agents an advantage and, thus, cement latent inequalities, but that contribute to prevent further costly disputes (Przepiorka et al. 2020). Gambetta (2009) describes a conflict-based formation of informal hierarchies among prisoners. The author explains why conflicts are more likely to occur in high-security prisons where inmates cannot move freely than in less heavily guarded prisons. Of course, one obvious explanation is that high-security prisons contain more aggressive inmates. However, Gambetta's analysis and the conflict dilemma described point to an alternative explanation: Information flow is restricted in maximum security prisons and conflicts must be staged to generate the missing information.

The section after the next one discusses in more detail how agents' limited knowledge about their own and other agents' preferences and constraints can be accounted for in game-theoretic models, and how this enhances our understanding of social dilemmas. But first it is discussed how game theory can be used to model agents' repeated encounters in social dilemmas.

5. REPETITION

The previous sections have occasionally alluded to how social dilemmas are solved or mitigated if agents encounter each other in the same social dilemma repeatedly. This section discusses this idea in more detail. It discusses the conditions under which repetition creates incentives for agents to cooperate in a cooperation dilemma like the PD shown in Figure 22.1a (with $T > R > P > S$). First, note that if the PD is repeated a predetermined number of times t, nothing changes in terms of agents' incentives to choose D in every round. The reasoning that leads to this conclusion was applied when the Nash equilibria in the sequential PD was derived (which is not the same as the repeated PD) and the TG (see above). Accordingly, using backward induction, one starts with the PD in the last round t, in which the two agents decide simultaneously and independently of each other whether to choose C or D. Since this is the last encounter, the interaction is equivalent to a one-shot PD, in which both agents choose D and get a payoff of P each. Since both agents anticipate this to be the outcome in the last round, they

Wojtek Przepiorka - 9781789906851

have no reason to behave differently in the second-to-last round ($t - 1$). Why would they? This reasoning continues back to the first round, in which the two agents come to the conclusion that choosing D is the best they can do. Hence, in the finitely repeated PD, mutual defection in every round remains the prevalent course of action.

This result seems counterintuitive because there are many instance in everyday life in which agents cooperate in finitely repeated interactions. Note, however, that in everyday life other factors come into play that promote cooperation. For example, a finitely repeated interaction may be embedded in other social relations (see Buskens and Raub 2013) or agents may have other regarding preferences that give them an intrinsic benefit from cooperating even in the last round (see next section). However, without relaxing the assumption that agents are rational and self-regarding (i.e. only care about their own payoffs), it can be shown how cooperation can emerge in the infinitely repeated PD under certain conditions.

The infinitely repeated PD, despite its name, is not only suitable to model-repeated interactions between immortal agents. The term 'infinitely' merely means that there is a certain probability δ (where $0 < \delta < 1$), with which the PD is repeated. Why would the agents' incentives to choose D change if they know they will encounter each other again with a certain probability, whereas their incentives do not change if they know how many times they will meet for sure? First, in the infinitely repeated PD, there is no predetermined last round and, therefore, backward induction cannot be applied. Second, if both agents choose C, they get a payoff of R each, which is larger than the payoff P they get from mutual defection. Although this is also true in the one-shot PD, only in the infinitely repeated PD do they have a possibility to enforce mutual cooperation. Mutual cooperation is enforced by the agents' choice of their strategies in the repeated game. Predicting what kind of strategies agents will choose from considering the Nash equilibria in infinitely repeated games is a thorny issue in game theory because there are so many possibilities. This issue is not expanded on here; it is shown, instead, how the choice of a particular equilibrium strategy by both agents creates an incentive for both to cooperate in every round.

While in the one-shot PD, agents' strategies are C or D (i.e. agents' strategies correspond to the actions they can take), in the repeated PD a strategy is a contingency plan that specifies an agent's actions in every round, dependent on what the other agent did in the previous rounds. Such strategies can be as simple as "Choose D in every round, irrespective of what the other agent did in the previous round" (All D); or "Choose C in every round, irrespective of what the other agent did in the previous round" (All C); and, more sophisticated, "Choose C in the first round, and in all subsequent rounds ($t > 1$) choose what the other agent chose in the previous round ($t - 1$)" (Tit-for-Tat). In his seminal book, *The Evolution of Cooperation*, Axelrod (1984) describes a computer tournament in which different strategies compete with each other in the repeated PD. The strategies were devised by researchers from different disciplines, programmed as algorithms that interacted with each other in what can be considered one of the first agent-based simulations. The winner of this tournament was the Tit-for-Tat strategy, which was submitted by mathematician and psychologist Anatol Rapoport (Axelrod 1984). The following paragraphs describe the conditions under which cooperation in the infinitely repeated PD can emerge by focusing on "Grim Trigger," another strategy that took part in Axelrod's computer tournament (see also Osborne 2009, chapter 14).

An agent using Grim Trigger chooses C in the first round and starts choosing D in all subsequent rounds following the round in which the other agent chooses D for the first time. In other words, Grim Trigger is a cooperative but unforgiving strategy. One could think of an entrepreneur

Wojtek Przepiorka - 9781789906851

that embarks on a joint venture with another entrepreneur determined to exit the project on the first occurrence of a defection by their partner. First, the outcome of an interaction between an agent 1 that uses All D and an agent 2 that uses Grim Trigger in the infinitely repeated PD is considered. Since agent 1 always defects and agent 2 starts defecting as from the second round, both agents' payoffs are the same as from the second round; they both get P in every round. In the first round, however, since agent 1 defects and agent 2 cooperates, agent 1 gets T and agent 2 gets S. Hence, agent 1 outperforms agent 2 in terms of payoffs since $T > S$. The expected payoff stream of agent 1 is:

$$T + \delta P + \delta^2 P + \delta^3 P + \ldots = T + \frac{\delta P}{1-\delta}. \tag{22.1}$$

The left-hand side of Equation (22.1) can be written as $T + \delta(P + \delta P + \delta^2 P + \ldots)$. The expression in parentheses is an (infinite) geometric series, and since $0 < \delta < 1$, the sum of the series is $P/(1 - \delta)$. Next, it is considered how an agent 1 that uses All C performs against an agent 2 that uses Grim Trigger. Since agent 1 always cooperates, agent 2 never starts defecting and both get R in every round from the start. Now the expected payoff stream of agent 1 (and agent 2) is

$$R + \delta R + \delta^2 R + \delta^3 R + \ldots = \frac{R}{1-\delta} \tag{22.2}$$

Given that agent 1, by choosing All D, outperforms agent 2, who chooses Grim Trigger as their strategy, why would agent 1 choose All C instead? Agent 1 would choose All C rather than All D only if the expected payoff stream from choosing All C (Equation 22.2) is larger than their expected payoff stream from choosing All D (Equation 22.1). That is, if:

$$\frac{R}{1-\delta} > T + \frac{\delta P}{1-\delta} \tag{22.3}$$

Since the payoffs of the PD are fixed, Equation 22.3 can be solved for δ to obtain

$$\delta > \frac{T-R}{T-P} \tag{22.4}$$

In other words, for agent 1 to choose All C rather than All D, the probability of encountering agent 2 again must be larger than a threshold value that is determined by the payoffs of the stage game. If the payoff variables are substituted for numbers typically used in the PD, for example $T = 3$, $R = 2$, $P = 1$ and $S = 0$, then agent 1 prefers All C over All D if $\delta > 0.5$. However, choosing All C as their strategy is not the best agent 1 can do because agent 2 could exploit agent 1 by choosing All D. If agent 1 chooses All C and agent 2 chooses All D, their expected payoff streams are $S/(1 - \delta)$ and $T/(1 - \delta)$, respectively. To avoid being exploited, agent 1 could choose All D as well, and both agents choosing All D establishes a Nash equilibrium in the infinitely repeated PD (neither agent can do better by unilaterally switching to another strategy). However, the All-D equilibrium is Pareto-inefficient. In fact, the All-D equilibrium is the best the two agents can do in the worst case because they both get a payoff stream of $P/(1 - \delta)$ (the *minimax* payoff), but cannot enforce cooperation. If both agents choose Grim Trigger instead, they both get a payoff stream of $R/(1 - \delta)$ and cannot do better by unilaterally switching to All D (or any other strategy). Since choosing D in any one round (e.g. as part of another strategy) would set them back to the minimax payoff (the other agent still uses Grim

Wojtek Przepiorka - 9781789906851

Trigger), by choosing Grim Trigger both agents establish a credible punishment threat that enforces cooperation. Hence, the Grim-Trigger equilibrium is one Nash equilibrium in the infinitely repeated PD in which cooperation prevails.

6. INFORMATION

The section on Cooperation introduced the trust dilemma (TD) as a model of social and economic exchange with trust at stake (see Figure 22.1d). It has been argued that, if not backed by social, institutional or organizational assurances, two agents encountering each other in a TD would not engage in an exchange. Moreover, it has been assumed that agents are perfectly informed about each other's preferences and constraints in terms of the payoffs of the stage game and other aspects of the interaction, such as the probability δ of encountering each other again. However, in many social dilemmas occurring in everyday life, agents hold private information about their preferences and constraints; they differ in their preferences and constraints and have information about their preferences and constraints that their interaction partners do not have. If, in a game with two agents, one agent holds private information about their own preferences and constraints and is informed about the other agent's preferences and constrains, the game is one with incomplete and asymmetric information (see Rasmusen 2007). The following paragraphs introduce the TD with incomplete and asymmetric information to illustrate how agents decide in a cooperation dilemma, if they have limited information about other agents' preferences and constrains (see also Raub 2004).

Figure 22.5 shows the TD with incomplete and asymmetric information (TDI). First note that the subgame starting with the right decision node of agent 1 corresponds to the TD shown in Figure 22.1d. Recall that in the Nash equilibrium of the TD, agent 2 chooses D because $T > R$ and agent 1 chooses D as well because $P > S$. In other words, since agent 2 is untrustworthy they are not trusted by agent 1. The subgame starting with the left decision node of agent 1 differs from the TD in that agent 2 has additional benefits and/or costs from choosing C or D, respectively, such that agent 2 prefers choosing C (i.e. is trustworthy). That is, for agent 2 it holds that $R + b > T - c$, where $b > 0$ and/or $c > 0$ (it will be discussed shortly what b and c may comprise substantially). The dashed line connecting the two decision nodes of agent 1 means that agent 1, when deciding whether to choose C or D, does not know whether they are at the left or the right decision node. In other words, agent 1 does not know whether they are about to interact with a trustworthy or

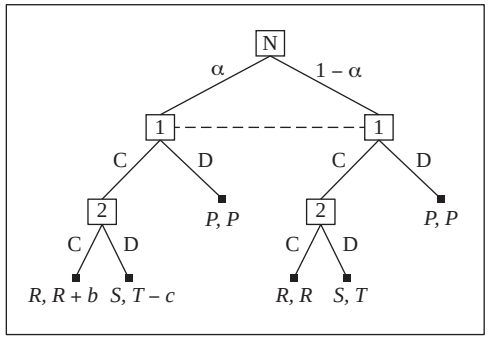


Figure 22.5 Trust dilemma with incomplete information

Wojtek Przepiorka - 9781789906851

untrustworthy agent 2. Agent 1 only knows that they are on either side with a certain probability. With probability α, agent 1 is about to interact with a trustworthy agent 2, and with probability $1 - \alpha$, agent 1 is about to interact with an untrustworthy agent 2. Whether agent 1 interacts with a trustworthy or untrustworthy agent 2 is decided by a move by Nature (N) that only agent 2 observes. Given the limited information agent 1 has about agent 2, how will agent 1 decide?

Agent 1 decides based on what they expect to get in either case. If agent 1 chooses C, their expected payoff is $\alpha R + (1 - \alpha)S$. If agent 1 chooses D, their payoff is P in either case. Agent 1 chooses C if choosing C gets them a larger expected payoff than choosing D. That is, if $\alpha R + (1 - \alpha)S > P$. Since the payoffs of the TDI are fixed, this equation can be solved for α to obtain:

$$\alpha > \frac{P - S}{R - S}. \tag{22.5}$$

In other words, agent 1 chooses C if the probability to interact with a trustworthy agent 2 is larger than a threshold value that is determined by the payoffs of agent 1. If the payoff variables are substituted for numbers typically used in the TD, for example $R = 2$, $P = 1$ and $S = 0$, then agent 1 chooses C if $\alpha > 0.5$. Of course, in $100 \times (1 - \alpha)$ -% of the cases, choosing C will get agent 1 the lowest payoff S, but as long as α is above the threshold specified in Equation 22.5, agent 1 does better (in expectation) by choosing C. But why are some agents trustworthy and others are not? Put differently, what do b and c in the TDI comprise?

In the TDI, b and c stand for the additional benefits and additional costs that agent 2 may have from choosing C or D, respectively. In an untrustworthy agent 2, these additional benefits and costs are absent (or not large enough to warrant that they choose C). These benefits and costs can be intrinsic (i.e. psychological) or extrinsic (i.e. contextual) (Riegelsberger et al. 2005). Intrinsic benefits and costs are also called social or other-regarding preferences (Kuhlman and Marshello 1975; Messick and McClintock 1968). For example, agent 2 may gain additional satisfaction ($+b$) from the gain that agent 1 makes in a cooperative exchange (Becker 1976) from the mere act of choosing C (Andreoni 1990), or agent 2 may be averse to feeling guilty ($-c$) as a consequence of their inflicting a loss on agent 1 by choosing D (Snijders and Keren 2001; Fehr and Schmidt 1999). Agents' extrinsic benefits and costs result from these agents' social and institutional embeddedness (Buskens and Raub 2013). For example, agent 2, by choosing D, may forgo the benefits from future exchanges ($+b$) that would accrue in a long-term exchange relation (Przepiorka and Diekmann 2013; Raub et al. 2019), suffer similar consequences from losing their good reputation (Shapiro 1983; Przepiorka 2013) or be negatively sanctioned by their peers or law enforcement ($-c$) (Diekmann and Przepiorka 2015; Gambetta and Przepiorka 2019).

7. SUMMARY

Social dilemmas are situations in which several agents face a decision conflict because of their interdependence. If everyone pursues their individual interests, the result can be a situation that puts everyone in a worse position than necessary. Social dilemmas occur in many situations in life, in interactions between family members, neighbors, work groups, companies and countries. They describe real and sociologically relevant phenomena and are thus an essential block in sociological theory building (see e.g. Coleman 1990).

Wojtek Przepiorka - 9781789906851

Game theory is a suitable tool to precisely describe social dilemmas. If one starts from a real-world social dilemma, the game-theoretic description of the situation forces one to think about the agents involved, their preferences, decision alternatives and the consequences of possible decision combinations. Although following such a procedure is lengthy and requires a certain degree of precision, it can bring counterintuitive relationships to light and prevent wrong conclusions (Watts 2014). Perhaps, most importantly for empirical-analytical sociology, game-theoretic models of social dilemmas can be used to derive hypotheses that can then be tested empirically. In this respect, game-theoretic models also provide indications of the necessary research designs and aspects of measurement of the variables mentioned in the hypotheses (Camerer 2003).

The aim of this chapter was to catch the reader's interest in game-theoretic modelling of social interactions. The chapter showed how game theory can be used to describe interaction situations to which cooperation problems, coordination problems and conflicts of interest are inherent. The chapter discussed how social, legal and organizational mechanisms can help to mitigate or resolve real social dilemmas, and it covered important elements of game-theoretic modelling of social dilemmas, such as repeated games and games with incomplete information, albeit by example rather than systematically. Needless to say, the brevity of this chapter does not allow us to go in depth on all of these topics or take up other ones, such as signaling games or models with agents with other-regarding preferences. The reader is, therefore, advised to also consider further readings.

REFERENCES

Andreoni, J. (1990), "Impure altruism and donations to public goods: A theory of warm-glow giving", *Economic Journal*, **100** (401), 464–77.

Axelrod, R. (1984), *The Evolution of Cooperation*, New York: Basic Books.

Becker, G.S. (1976), "Altruism, egoism and genetic fitness: Economics and sociobiology", *Journal of Economic Literature*, **14** (3), 817–26.

Bigoni, M., S.-O. Fridolfsson, C. Le Coq and G. Spagnolo (2015), "Trust, leniency, and deterrence", *Journal of Law, Economics, and Organization*, **31**: 663–89.

Buskens, V. and W. Raub (2013), "Rational choice research on social dilemmas: Embeddedness effects on trust", in R. Wittek, T. Snijders and V. Nee (eds.), *The Handbook of Rational Choice Social Research*, pp. 113–50, Stanford, CA: Stanford University Press.

Camerer, C.F. (2003), *Behavioral Game Theory*, Princeton, NJ: Princeton University Press.

Coleman, J.S. (1990), *Foundations of Social Theory*, Cambridge, MA: The Belknap Press of Harvard University Press.

Darley, J.M. and B. Latané (1968), "Bystander intervention in emergencies: Diffusion of responsibility", *Journal of Personality and Social Psychology*, **8** (4): 377–83.

Dasgupta, P. (1988), "Trust as a Commodity", in D. Gambetta (ed.), *Trust: Making and Breaking Cooperative Relations*, pp. 49–72, Oxford: Basil Blackwell.

Dawes, R.M. (1980), "Social dilemmas", *Annual Review of Psychology*, **31**: 169–93.

Diekmann, A. (1985), "Volunteer's dilemma", *Journal of Conflict Resolution*, **29** (4): 605–10.

Diekmann, A. (1993), "Cooperation in an asymmetric volunteer's dilemma game: Theory and experimental evidence", *International Journal of Game Theory*, **22** (1): 75–85.

Diekmann, A., B. Jann and D. Wyder (2009), "Trust and Reputation in Internet Auctions", in K.S. Cook, C. Snijders, V. Buskens and C. Cheshire (eds.), *eTrust: Forming Relationships in the Online World*, pp. 139–65, New York: Russell Sage.

Diekmann, A. and W. Przepiorka (2015), "Punitive preferences, monetary incentives and tacit coordination in the punishment of defectors promote cooperation in humans", *Scientific Reports*, **5**, 10321.

Diekmann, A. and W. Przepiorka (2016), "'Take one for the team!' Individual heterogeneity and the emergence of latent norms in a volunteer's dilemma", *Social Forces*, **93** (3): 1309–33.

Diekmann, A. and W. Przepiorka (2019), "Trust and Reputation in Markets", in F. Giardini and R. Wittek (eds.), *The Oxford Handbook of Gossip and Reputation*, pp. 383–400, Oxford: Oxford University Press.

Wojtek Przepiorka - 9781789906851

Fehr, E. and S. Gächter (2002), "Altruistic punishment in humans", *Nature*, **415** (6868), 137–40.
Fehr, E. and K.M. Schmidt (1999), "A theory of fairness, competition, and cooperation", *Quarterly Journal of Economics*, **114** (3), 817–68.
Gambetta, D. (2009), *Codes of the Underworld: How Criminals Communicate*, Princeton, NJ: Princeton University Press.
Gambetta, D. and W. Przepiorka (2019), "Sharing compromising information as a cooperative strategy", *Sociological Science*, **6**: 352–79.
Garrett, G. and B.R. Weingast (1993), "Ideas, Interests, and Institutions: Constructing the European Community's Internal Market", in J. Goldstein and R.O. Keohane (eds.), *Ideas and Foreign Policy: Beliefs, Institutions, and Political Change*, pp. 173–206, Ithaca, NY: Cornell University Press.
Gintis, H. (2009), *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction* (2nd edn), Princeton, NJ: Princeton University Press.
Granovetter, M. (1992), "Problems of Explanation in Economic Sociology", in N. Nohria and R.G. Eccles (eds), *Networks and Organizations: Structure, Form, and Action*, pp. 25–56, Boston, MA: Harvard Business School Press.
Greene, J. (2013), *Moral Tribes: Emotion, Reason and the Gap Between Us and Them*, New York: Penguin Books.
Güth, W. and A. Ockenfels (2003), "The Coevolution of Trust and Institutions in Anonymous and Non-Anonymous Communities", in M.J. Holler, H. Kliemt, D. Schmidtchen and M. Streit (eds), *Jahrbuch für Neue Politische Ökonomie*, pp. 157–74, Tübingen: Mohr Siebeck.
Hardin, R. (2002), *Trust and Trustworthiness*, New York: Russell Sage Foundation.
Harsanyi, J.C. and R. Selten (1988), *A General Theory of Equilibrium Selection in Games*, Cambridge, MA: MIT Press.
Heckathorn, D.D. (1989), "Collective action and the second-order free-rider problem", *Rationality and Society*, **1** (1): 78–100.
Hedström, P. and P. Bearman (eds) (2009), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press.
Kollock, P. (1998), "Social dilemmas: The anatomy of cooperation", *Annual Review of Sociology*, **24**: 183–214.
Kreps, D. (1990), "Corporate Culture and Economic Theory", in J.E. Alt and K.A. Shepsle (eds), *Perspectives on Positive Political Economy*, pp. 90–143, Cambridge, MA: Cambridge University Press.
Kuhlman, D.M. and A. Marshello (1975), "Individual differences in the game motives of own, relative, and joint gain", *Journal of Research in Personality*, **9** (3), 240–51.
Lau, S.-H.P. and V.-L. Mui (2012), "Using turn taking to achieve intertemporal cooperation and symmetry in infinitely repeated 2 × 2 games", *Theory and Decision*, **72** (2): 167–88.
Manzo, Gianluca (ed.) (2014), *Analytical Sociology: Actions and Networks*, Chichester: Wiley.
McAdams, R.H. (2009), "Beyond the prisoners' dilemma: Coordination, game theory, and law", *Southern California Law Review*, **82** (2): 209–58.
Messick, D.M. and C.G. McClintock (1968), "Motivational bases of choice in experimental games", *Journal of Experimental Social Psychology*, **4** (1), 1–25.
Olson, M. (1965), *The Logic of Collective Action: Public Goods and the Theory of Groups*, Cambridge, MA: Harvard University Press.
Osborne, M.J. (2009), *An Introduction to Game Theory, International Edition*, Oxford: Oxford University Press.
Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge, UK: Cambridge University Press.
Poundstone, W. (1992), *Prisoner's Dilemma*, New York: Anchor Books.
Przepiorka, W. (2013), "Buyers pay for and sellers invest in a good reputation: More evidence from eBay", *Journal of Socio-Economics*, **42** (C): 31–42.
Przepiorka, W. (2020), "Soziale Dilemmas", in A. Tutic (ed.), *Rational-Choice-Theorie*, pp. 181–99, Berlin: De Gruyter Oldenbourg.
Przepiorka, W. and J. Berger (2016), "The sanctioning dilemma: A quasi-experiment on social norm enforcement in the train", *European Sociological Review*, **32** (3): 439–51.
Przepiorka, W. and A. Diekmann (2013), "Temporal embeddedness and signals of trustworthiness: Experimental tests of a game theoretic model in the United Kingdom, Russia, and Switzerland", *European Sociological Review*, **29** (5), 1010–23.
Przepiorka, W. and A. Diekmann (2018), "Heterogeneous groups overcome the diffusion of responsibility problem in social norm enforcement", *PLOS ONE*, **13** (11): e0208129.
Przepiorka, W., L. Norbutas and R. Corten (2017), "Order without law: Reputation promotes cooperation in a cryptomarket for illegal drugs", *European Sociological Review*, **33** (6): 752–64.
Przepiorka, W., C. Rutten, V. Buskens and A. Szekely, (2020), "How dominance hierarchies emerge from conflict: A game theoretic model and experimental evidence", *Social Science Research*, **86**: 102393.
Rapoport, A. (1974), "Prisoner's Dilemma – Recollections and Observations", in A. Rapoport (ed.), *Game Theory as a Theory of Conflict Resolution*, pp. 17–34, Dordrecht: Reidel.
Rapoport, A. and A.M. Chammah (1966), "The game of chicken", *American Behavioral Scientist*, **10** (3): 10–28.

Wojtek Przepiorka - 9781789906851

Rasmusen, E. (2007), *Games and Information: An Introduction to Game Theory*, Malden, MA: Blackwell Publishing.
Raub, W. (2004), "Hostage posting as a mechanism of trust: Binding, compensation, and signaling, *Rationality and Society*, **16** (3), 319–65.
Raub, W., V. Buskens and R. Corten (2015), "Social Dilemmas and Cooperation", in N. Braun and N.J. Saam (eds), *Handbuch Modellbildung und Simulation in den Sozialwissenschaften*, pp. 597–626, Wiesbaden: Springer VS.
Raub, W., V. Buskens and V. Frey (2019), "Strategic tie formation for long-term exchange relations", *Rationality and Society*, **31** (4), 490–510.
Riegelsberger, J., M.A. Sasse and J.D. McCarthy (2005), "The mechanics of trust: A framework for research and design", *International Journal of Human-Computer Studies*, **62** (3), 381–422.
Schelling, T.C. (1960), *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.
Selten, R. (1975), "Reexamination of the perfectness concept for equilibrium points in extensive games", *International Journal of Game Theory*, **4** (1): 25–55.
Shapiro, C. (1983), "Premiums for high quality products as return to reputation", *Quarterly Journal of Economics*, **98** (4): 659–80.
Snijders, C. and G. Keren (2001), "Do you trust? Whom do you trust? When do you trust?", *Advances in Group Processes*, **18**: 129–60.
Sosis, R. (2005), "Does religion promote trust? The role of signaling, reputation, and punishment", *Interdisciplinary Journal of Research on Religion*, **1**: 1–30.
Van Lange, P.A.M., J. Joireman, C.D. Parks and E. van Dijk (2013), "The psychology of social dilemmas: A review", *Organizational Behavior and Human Decision Processes*, **120** (2): 125–41.
Voss, T. (2001), "Game-Theoretical Perspectives on the Emergence of Social Norms", in M. Hechter and K.-D. Opp (eds), *Social Norms*, pp. 105–36, New York: Russell Sage Foundation.
Watts, D.J. (2014), "Common sense and sociological explanations", *American Journal of Sociology*, **120** (2): 313–51.
Wikipedia Contributors (2018), Soziales Dilemma. Wikipedia, Die freie Enzyklopädie. Retrieved on October 8, 2019 at https://de.wikipedia.org/wiki/Soziales_Dilemma.
Yamagishi, T. (1986), "The provision of a sanctioning system as a public good", *Journal of Personality and Social Psychology*, **51** (1): 110–16.

Wojtek Przepiorka - 9781789906851

23. Network mechanisms and network models

Christoph Stadtfeld and Viviana Amati

1. INTRODUCTION

Societies can be understood as complex systems of interdependent social actors that are tied to one another in various ways. While the individual actors may have a limited understanding of the system as a whole and mostly respond to their local environments (the micro level), their interdependent actions shape the macro-level features that in turn affect the opportunities of actors to act and interact. The micro-macro motif is a central concept in analytical sociology (Hedström and Bearman 2009; Coleman 1990) and is discussed in several chapters in this book (see in particular Chapters 1, 2 and 8).

Social networks are a formal framework to describe, analyze, model, and theorize about these interdependent social systems. Beyond their metaphorical use, social networks have an underlying mathematical language. In its most basic form, we may denote a social network by a set of actors (the nodes) and a set of binary relations (the ties). This basic formal definition of a network may be extended by considering that often multiple networks exist in the same social context, that individuals can be tied to other types of entities, that networks change through time, and that they co-evolve with individual attributes.

Social networks are a promising framework to study micro-macro links in interdependent social systems (Stadtfeld 2018). Individuals' positions in networks (e.g. their degree centralities, embeddedness, similarity) can be formalized using social network concepts. These positions may be viewed as micro-level features that characterize individuals' social environments that matter for their actions. Interpersonal social networks can also be regarded as the meso level in the micro-macro process, for example, when they are viewed as conduits of interpersonal social influence that explain macro-phenomena such as diffusion of innovation (Roux and Manzo 2018; Manzo 2007).

Finally, we can consider *complex network concepts* as a representation of macro-level structures and processes. Those complex representations can include multiple layers of interpersonal networks (e.g. interaction, cooperation, trust), their interdependent change, affiliation structures represented as two-mode networks, and individual attributes. Complex network representations can be utilized to formalize macro-level outcomes and processes that relate to connectedness and segregation, status hierarchies, affiliations and access to resources, or the distribution of individual variables such as knowledge and economic capital. Utilizing network concepts on different levels enables a close link between explanatory theory, empirical data, and statistical modeling in sociological micro-macro studies.

This chapter introduces strategies for the analytical investigation of social mechanisms in social networks. The goal is to explain structural properties and change in social networks from individual behavior that is affected by their local network positions. We first introduce the concept of *network mechanisms* and compare and contrast it with the notion of social mechanisms in analytical sociology (Hedström and Swedberg 1998). We then focus on how those network mechanisms are tested with *network models* in a variety of empirical settings

Christoph Stadfeld and Viviana Amati - 9781789906851

and with different types of network data. We further discuss how some of these statistical models may be applied as agent-based simulation models to understand the role of network mechanisms in emerging network properties.

1.1 Network Mechanisms

The goal of network models is to explain the presence of ties in social networks, or their change through time. In this section, we define and discuss the concept of *network mechanisms*—structural processes that are regularly found in social networks and that can be linked to one or multiple causal *social mechanisms* of network tie formation.

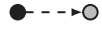
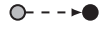
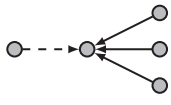
Social mechanisms are a core concept of analytical sociology (Hedström and Swedberg 1998). Social theory in this view can be dissected into causal mechanisms that explain how a given input and output are connected through a sequence of events (Manzo 2014). The output of the mechanism is the social phenomenon to be explained. Manzo states that "a mechanism can be conceived as consisting of a set of organized entities whose properties and activities are able to trigger changes that generate the observed connections with some regularity" (p. 14). Manzo, as well as Hedström and Ylikoski (2014), argue that a generic definition of social mechanism is hard to provide, as the specifics are inseparable from the output being studied and the level (micro, meso, macro) on which it operates. However, Hedström and Ylikoski (2014: p. 62) posit that social mechanisms share certain characteristics: they are defined as *causal* chains of events, they are explicit about their internal *structure* in terms of entities and their relations, and they form a *hierarchy* where each mechanism can be further dissected into sub-mechanisms in which specific properties or activities of the entities are explained on a lower level in more detail.

We define network mechanisms as classes of multiple social mechanisms that occur in the same structural positions (the input of the mechanism) and explain the presence of a tie in a network (the output). We argue that a single network mechanism is typically not sufficiently specific to be considered a social mechanism in the spirit of analytical sociology as it typically cannot be linked to one single causal explanation. However, when multiple network mechanisms are combined in one network model, or when their formulation is further tailored to a specific causal explanation, they can be considered as social mechanisms. Network mechanisms in our definition are explicit about the internal structure of entities and relations, and can naturally be embedded in a hierarchy. Generic network mechanisms (see Table 23.1) serve as starting points that can be either refined to become more specific, or be combined into sequences of change events to represent more complex types of mechanisms.

The *structural position* of a network mechanism is specified by considering network ties and actors' attributes in the environment of the explained tie. In directed networks, we refer to the two actors as the sender and receiver of the tie. For example, a structural position can be specified by assessing whether sender and receiver are indirectly connected through a third actor, or whether they are central actors in the network under study. Similarly, actor and dyad-level attributes can be used to specify a network position. Actor attributes can include demographic variables, behaviors, beliefs, or preferences, while dyad-level attributes may include social and spatial distances or a tie in another network. For example, a structural position can specify whether two actors have similar attributes or are close in a social space. By defining structural positions as a function of network structures and attributes, structural positions relate to the generic concept of *network positions* that was introduced to characterize positions of nodes (rather than ties) in social networks (Brandes 2016).

Christoph Stadfeld and Viviana Amati - 9781789906851

Table 23.1 Most common network mechanisms and their representation. Dashed ties are the explained ties. Continuous ties and black nodes represent elements defining the structural position of the network mechanism

Network mechanism	Representation	Network mechanism	Representation
Reciprocity		Activity	
Transitivity		Attraction	
Popularity		Homophily	

The explanantia (model inputs) of network models are thus flexible formulations of structural positions of actors. The explanandum (model output) are the ties between actors—such as their interaction, cooperation, social support, or friendships. In principle, actor attributes can also be considered the output of a network model (for example, to explain processes of social influence, diffusion, and polarization in networks) leading to a similar specification of mechanisms that is not further considered in this chapter.

The most basic and widely used network mechanisms are depicted in Table 23.1 and some causal explanations for their presence are discussed in the following.

Reciprocity expresses the tendency to establish a tie with others who are tied to oneself. The structural position is formally defined by the presence of a tie from receiver to sender of the tie to be explained. Theoretical arguments for why reciprocal ties emerge are, for example, based on social exchange theory (Emerson 1976), or social and spatial proximity of actors in a network.

Transitivity is one of the central mechanisms that goes beyond the dyad and thus involves more than two actors: it expresses the tendency of actors to connect to others to whom they are indirectly tied to through a third actor. The structural position is defined by the presence of a "two-path" between the sender and the receiver of the explained tie. Multiple causal explanations exist that predict transitive closure in a social network. Among those are differences in spatial or social distances and similarity attraction based on an actor attribute (Granovetter 1973). One prominent theoretical explanation is Heider's Balance Theory (Heider 1958) that suggests that actors may perceive cognitive dissonance and stress if they are not in a positive relationship with those they are indirectly positively tied to.

Popularity is the tendency of actors to connect to those who receive ties from many others. The structural position is defined by the in-degree of the receiver of an explained tie. This mechanism is, for example, in line with the Matthew effect mechanism in science discussed by Merton (1968). Increasing recognition of an actor's scientific work (e.g. number of ties in a citation network) may be a signal that amplifies the beliefs of other scientists about this scientist's importance, which then causes them to cite the actor with a higher likelihood, further increasing its in-degree. Similar re-enforcing mechanisms have been labeled *preferential attachment* in

Christoph Stadfeld and Viviana Amati - 9781789906851

the work of Barabási and Albert (1999). But the presence of a popularity mechanism is not sufficient evidence that a Matthew effect exists in a network. Other social mechanisms may explain a similar disproportional growth of network ties. For example, other actor attributes, such as the career stage, might send similar signals that affect the beliefs of others.

Activity and *Attraction* are the tendencies that actors with specific attributes will be more likely to send and receive ties, respectively. The structural positions are defined by attributes of the sender and receiver of the explained tie, respectively. Whether these differences are, for example, rooted in actors' beliefs about actors with a specific attribute (in case of Attraction) or whether they relate to differences in networking resources (in case of Activity) is not further specified by the mechanism.

Homophily expresses the tendency that similar actors are more likely to connect (Lazarsfeld, et al. 1954; McPherson et al. 2001). Homophily is, as in the previous examples of mechanisms, no causal explanation in itself, but a superordinate network mechanism that subsumes multiple social mechanisms. Its structural position is defined by the attribute similarity of the sender and receiver of the explained tie. Various causal explanations exist that can lead to this outcome. These may be cognitive processes about similarity attraction (Huston and Levinger 1978) or structural processes that are affected by existing baseline segregation and social distances in the social setting under study.

We highlighted that none of the presented network mechanisms relate straightforwardly to one single causal mechanism in the spirit of Hedström and Swedberg (1998). Instead, they are broad and widely observed classes that subsume a number of social mechanisms that connect a specific input (e.g. being similar, being indirectly tied) to the outcome of the model, the observation of a network tie.

To understand the causal social mechanisms that connect input and output, and control for reinforcement, we can take two directions. First, we can simultaneously include several network mechanisms in one *multi-mechanistic network model*. This enables a residual interpretation of a single network mechanism that helps to narrow down plausible causal explanations. If, for example, we find evidence for the transitivity mechanism while testing for homophily on a specific attribute, we may conclude that the residual effect of transitivity is not merely explained by similarity attraction, but must be the consequence of another causal explanation, e.g. structural balance, social distances, or similarity on a different attribute. We can further narrow down the set of possible explanations by including additional alternative network mechanisms. By including several network mechanisms in one model we thus arrive at more specific interpretations.

Second, we may refine the definition of structural positions of the network mechanisms to be a better reflection of the underlying social mechanism. In this case, the level of specificity should align with the specificity of the theoretical explanation of interest. If, for example, one is interested in testing social mechanisms about stress-induced closure of triadic structures as predicted by structural balance theory, one could formulate network mechanisms that explicitly include measurements of situational stress in imbalanced interactions. The structural position of such a mechanism could be defined by the situational level of stress of actors when they encountered social situations in which both a friend and a friend's friend (tertius) participated, while they do not see themselves as friends with tertius, yet. Such a fine-grained definition of a structural position, however, implies that data and models should incorporate the necessary situational and perception measurements, and their relative timing. It should further be noted that fine-grained network mechanisms necessarily increase model complexity.

Christoph Stadfeld and Viviana Amati - 9781789906851

Multi-mechanistic interpretation and increasing specificity of the defined network mechanisms are two important ways of how empirical research can approach causal explanations of social phenomena. This refinement of network mechanisms is in line with the idea of the hierarchical organization of social mechanisms discussed by Hedström and Ylikoski (2014). The generic mechanisms presented form a first hierarchical level that can be dissected further to approach the notion of causality.

Social scientists need to carefully balance out the aim for simplicity of their models that facilitate *interpretability* while maximizing their specificity and *explanatory power*. The generic classes of network mechanisms discussed in this section are useful starting points for developing network models that aim to understand the micro-level causes of social structure.

1.2 Network Models

The field of social network analysis (SNA) has proposed a plethora of descriptive statistics to formally measure and compare network features such as the levels of clustering, similarity, reciprocity, and centrality. A good introduction to such descriptive SNA is, among many others, the work by Wasserman and Faust (1994). Here, we focus on inferential SNA and, in particular, on network models.

The term *network models* refers to a class of statistical models in which the dependent variable is the network and the independent variables are the network ties themselves and the node- and dyad-level variables. Formally, a network model is a probability distribution on networks defined on a set of entities. The richness of network models stems from the specification of the probability distribution.

This definition of a network model derives from the type of inference the network models are based on. Unlike classical social settings, in which the data are collected using a sampling design, often consisting of selecting a (large) sample of independent units from a population, in network analysis typically only one observation of the network is available. We assume that the observed network is one instance of a population of networks whose probability distribution is described by the model. The model is a formula accounting for the social mechanisms that might have generated the observed network and the randomness deriving from unobserved variables or elements not directly accounted for by the model.

There are several classes of network models that have been developed over the past decades. The first class of models has its roots in statistical mechanics and computer science. The models belonging to this class are probability distributions defining the network mechanisms that might have generated networks with specific network-level properties, such as degree distribution, clustering, and connectivity. Examples are small-world (Watts and Strogatz 1998) and preferential attachment models (Barabási and Albert 1999). Those models are applied to investigate how network-level features are generated from simple rules of tie formation, or are used as reference distribution in testing hypotheses concerning the network mechanisms. These models typically consider only a few specific mechanisms and do not aim at capturing the multi-mechanistic nature of empirical social networks. We refer to the work of Watts (2004) and the references therein for details on those stylized network models.

The second class of models stems from the interest of social scientists in the micro-level properties of networks and how local patterns of ties generated the structure of an observed network. Two model sub-classes belong to this class of models. The first includes models that control for the network structure, i.e. test the significance of a social process while controlling

Christoph Stadfeld and Viviana Amati - 9781789906851

for specific structural characteristics of the network (Wasserman and Faust et al. 1994). The second comprises the models that explicitly model tie dependence and test the relevance of a social mechanism while simultaneously considering other competing mechanisms (Robins et al. 2007; Snijders et al. 2010). Both sub-classes are applied to simulate social networks with partly realistic features and determine how the networks would look if a set of mechanisms had operated.

The third class of models comprises the latent variable models. This family includes latent position models (e.g. Hoff et al. 2002) and stochastic block-models (e.g. Holland et al. 1983) as special cases. The former associate a latent position in Euclidean space to each node and assume that nodes that are close in that space are more likely to be linked. A logistic regression model links tie probabilities to the distance among the nodes. The latter aims at clustering nodes into groups referred to as blocks and models the probability of ties between nodes as a function of the blocks.

In the following, we focus on the second class of models. We refer the readers to the works cited in the previous lines for details on the models belonging to the first and third class.

1.3 Network Data

Regular social science data are compositional and represent characteristics and behaviors of a set of actors $N = \{1, \ldots, n\}$.Those data are usually collected using sample designs in which the unit of analysis constitutes a sample of independent observations. Network data combine compositional data with relational data, providing information on the relations between actors. They are collected on pairs of actors $\{i, j\}$, referred to as dyads, and thus are a particular form of dyadic data (Brandes et al. 2013). Since dyads might share actors, the units of analysis in network studies are interdependent.

Several types of variables are collected over dyads. Tie variables indicate whether there is a relation between actors i and j. We denote a tie variable by X_{ij}. Although X_{ij} can take any real value, we restrict our attention to binary tie variables. The collection of the tie variables X_{ij} and the set of actors constitute the network. We represent a network by the $n \times n$ adjacency matrix $X = \{X_{ij,\, i,j \in N}\}$. We say that a network is directed if the existence of the tie from i to j does not imply the existence of a tie from j to i (e.g. giving advice), and undirected otherwise (e.g. co-authorship).

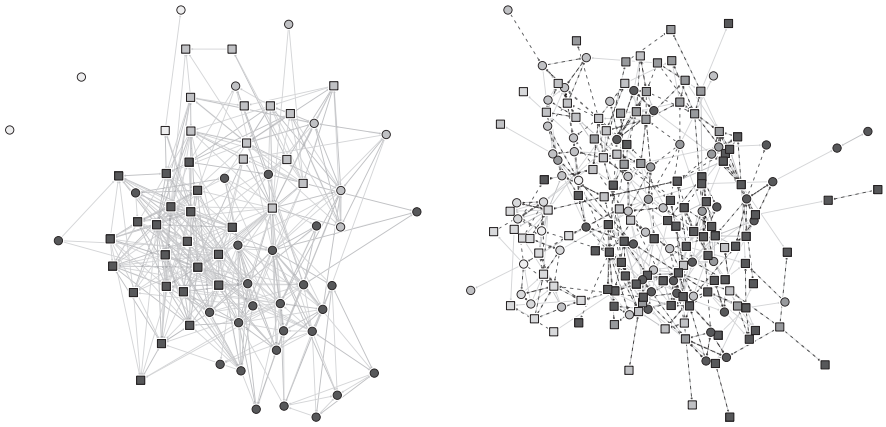
Other types of variables are compositional in nature and are measured at the actor or the dyadic level. We denote by v and w the node- and the dyad-level variables, respectively.

A network can be observed at one single time point in a cross-sectional design or over time in a longitudinal design. Dependence on time is denoted by the time index $t \in T$ so that X_t is the network at time t. The set T is discrete when the network is observed in panel designs and continuous when the observation of the network takes place at more fine-grained intervals. In the following, we use upper-case letters to denote random variables. Observed values of random variables or fixed quantities are represented by lower-case letters. Thus, for instance, X denotes the network random variable and x the observed network.

1.4 Empirical Examples

We illustrate the introduced models by using two datasets. The former concerns friendship and seeking advice networks among the 71 attorneys (36 partners and 35 associates) of a Northeastern

Christoph Stadfeld and Viviana Amati - 9781789906851



(a) Lazega's lawyers (b) Swiss StudentLife data

Figure 23.1 (a) Friendship network of Lazega's lawyers data. The shape of the nodes represents the attorneys' type (square = partner, circle = associate), and the colour of the office (light gray = Boston, gray = Hartford, dark gray = Providence). Advice relations are not shown as the network is very dense. (b) Friendship (solid line) and support network (dashed line) of the Swiss StudentLife study. The shape of the node represents the gender (square = male, circle = female) and the color of the study major

US corporate law firm during the period 1988–1991 (Lazega et al. 2001). The data set[1] includes information on a few entity levels attributes: seniority (measured as the number of years spent working for the firm), the office in which the lawyers work (Boston, Hartford, and Providence), and the school in which the lawyers studied (Harvard and Yale, UConn, and other universities).

A second data set is taken from the Swiss StudentLife study (Vörös et al. 2021). It consists of social relations collected in a community of 226 first-year students at a Swiss university (Stadtfeld et al. 2019). Network data and actor attributes were collected multiple times throughout their studies. In one example, we focus on friendship between and seeking informational support from other students, as these are similar in interpretation to the networks considered in the first study. In another example, we investigate data of face-to-face interactions collected at the beginning of the study year. The actor attributes considered are binary gender and specialization in different study programs.

Figures 23.1a and 23.1b show the friendship network of Lazega's lawyers data and an excerpt of the friendship and support network of the StudentLife study, respectively.

2. STATISTICAL NETWORK MODELS

2.1 Stylized Network Models: CUGs

The simplest network models are stylized representations. They are reference distributions against which we can test whether there is evidence for a network mechanism while controlling

[1] Full description and data available at https://www.stats.ox.ac.uk/~snijders/siena/

Christoph Stadfeld and Viviana Amati - 9781789906851

for a set of network properties. Those distributions are usually referred to as conditional uniform graphs (CUGs) because they assign a uniform probability to networks having the same desired features and a null probability to those not having that features.

The simplest CUG is obtained by conditioning on the network density. This model is known in the literature as the Erdős–Renyi–Gilbert model (Erdős and Rényi 1959; Gilbert 1959). More complex network properties can be conditioned on, for example, the level of reciprocation, the degree distribution of the network, or the dyad census. We refer to Wasserman and Faust (1994: chapter 13) for more details on these distributions.

To test hypotheses, researchers specify CUGs by conditioning on the features they want to control for, sample a large number of networks from the obtained model, and count the number of sub-graph structures representing the network mechanism in the generated networks. Counts of sub-graph structures are hereafter referred to as statistics. They resemble counts of structural positions as defined for network mechanisms, *after* the inclusion of the explained tie. The distribution of the statistics is the reference distribution against which the observed values of the statistics are compared. An empirical p-value is computed to test whether the observed statistics are more or less likely under the assumption that the network is from the reference distribution.

We illustrate the application of CUGs using Lazega's data. We test for reciprocity, transitivity, and school homophily against a CUG conditioning on the network density. Figure 23.2 shows the distributions of the three statistics obtained by sampling 5 000 networks from the CUG and the observed value (dotted lines). Compared to the observed, all the generated networks have a lower number of mutual dyads and transitive triangles. Thus, the empirical p-value is 0. For homophily, 7.8% of the generated networks have an equal or larger number of

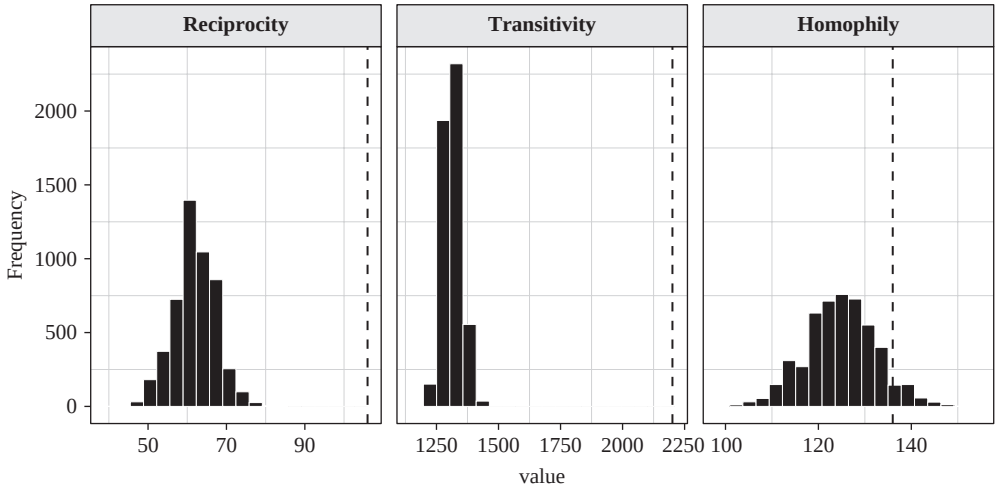


Figure 23.2 Distributions of mutual and homophilous dyads, and transitive triads obtained by sampling 5 000 networks from a CUG conditional on the observed number of advice ties of Lazega's data. The dashed line represents the observed values. The empirical p-value is the proportion of simulated values greater than or equal to the corresponding observed values

Christoph Stadfeld and Viviana Amati - 9781789906851

same-school ties, i.e. the empirical p-value is 0.078. Thus, conditioning on the network density, there is evidence for reciprocity, transitivity, but not for school homophily.

The application of CUGs has two main limitations. The former is a technical issue. The combinatorial complexity of CUGs increases with the number of conditioning features. Thus, the network characteristics that can be controlled for are in practice restricted to those described. The latter relates to the interpretation of the results of CUG tests. The results do not generalize to the phenomenon under study but are conditional on the considered features. Therefore, we can only infer whether the occurrence of a specific network characteristic in the observed network is more or less likely compared to its prevalence in a random graph with particular features.

2.2 Quadratic Assignment Procedure Regression

The quadratic assignment procedure regression (QAP regression) aims to assess whether entity and dyadic attributes and other relations among the entities might be predictive factors of the occurrence of ties in an observed network. Similar to CUGs, a QAP regression tests hypothesis on the existence of ties and does not model the dependence of ties directly. Compared with CUGs, hypothesis testing in QAP regression is conditional on the structure of the entire network (i.e. controls for the network structure), and multiple hypotheses can be tested simultaneously while controlling for one another.

Formally, the QAP regression takes the form of a linear model in which the dependent variables are the tie random variables X_{ij}, and the explanatory variables Z are structural positions of network mechanisms. These are specified by actor covariates of sender and receiver, dyadic covariates, or other relations among the actors. The general form of the model is

$$g(X_{ij}) = \beta_0 + \beta_1 Z_1 + \ldots + \beta_p Z_p,$$

with g the link function. When the network is valued, g is the identity function and the QAP regression takes the form of a standard linear regression model. When the network is binary, g is the logit function and the QAP regression takes the form of a logistic regression model.

Like in standard regression models, testing the model coefficients β allows making inference on the direction and strength of the association between the occurrence of ties and the explanatory variables. However, standard practices based on t- or z-statistics are not valid because network data do not meet the assumption of independent observations.

Quadratic assignment procedures apply to overcome this issue. They use permutations of the original data to compute a reference distribution of the parameter estimators and the empirical p-values. The procedure comprises two steps. In the first step, the network data are permuted many times in such a way that the structure of the network remains unaltered. Several permutation strategies are available. We refer to Krackhardt (1988) and Dekker et al. (2007) for the technical details. In the second step, the regression model is estimated on each permuted data, and the t(z)-values are computed. The distribution of these t(z)-values is the empirical distribution of the T(Z)-test statistic. The empirical p-value is the proportion of permutations leading to a t(z)-value (in absolute value) greater than or equal to the observed t(z)-value.

Christoph Stadfeld and Viviana Amati - 9781789906851

Table 23.2 QAP regression investigating the association between advice ties and seniority, office, and friendship. The first column reports the estimated coefficients, the second the odds ratios, and the last the empirical p-values computed using 5 000 permutations

	Est.	exp(Est.)	p-value	Sig.
Intercept	–1.80	0.17	<0.001	***
Seniority (receiver-sender)	0.04	1.04	<0.001	***
Friendship	1.85	6.36	<0.001	***
Same Office	0.84	2.32	<0.001	***
Same School	0.18	1.20	0.23	

Notes: **p*-value < 0.05, ***p*-value < 0.01, ****p*-value < 0.001.

For illustrative purposes, we use the QAP logistic regression to test whether seniority, working in the same office, having studied in the same school, and friendship ties are predictive factors of seeking advice ties for Lazega's data.

Table 23.2 reports the results based on 5 000 permutations of the original network. All the coefficients are significantly different from 0, but related to the same school. Advice ties from lawyers with low seniority to lawyers with high seniority (OR = 1.04, percentage change 4%), to friends (OR = 6.36, percentage change 536%), and between lawyers who attended the same office (OR = 2.32, percentage change 132%) are more likely. There is no evidence of an association between advice ties and the school in which the lawyers studied. This finding is coherent with the results of the CUG tests presented in the previous section. However, the interpretation is now conditioning on the empirically observed network structures, and thus more specific to networks of the observed type. From the CUG test, we conclude that there is no evidence that same-school ties are more likely to occur in the observed network than in networks having the same number of edges. From the QAP regression, we infer that there is no evidence of an association between the advice ties and the school in the observed network since we are controlling for the entire network structure and not only for its number of edges. With QAP regressions we can only meaningfully test network mechanisms in which the structural position is defined by nodal- and tie-level covariates.

2.3 Exponential Random Graph Models

Exponential Random Graph Models (ERGMs) provide a probability distribution over a population of binary graphs with a fixed number of nodes. ERGMs express the effect of node and tie attributes on the probability of a tie to be formed (e.g. activity, attraction, homophily) as known from QAP regressions, but at the same time they can include effects expressing the prevalence of network mechanisms (e.g. reciprocity, transitivity, popularity) that were merely conditional features in CUGs and QAP regression. ERGMs are *multi-mechanistic* network models because they consider multiple network mechanisms at the same time, allowing assessment of their relative importance in explaining the observed network data. Compared to QAP regression they not only control for the presence of structure in the network data, but they model it explicitly.

Christoph Stadfeld and Viviana Amati - 9781789906851

The ERGM probability is formally defined as

$$P(X = x) = \frac{\exp\left(\sum_k \beta_k s_k(x)\right)}{\sum_{x' \in \mathcal{X}} \exp\left(\sum_k \beta_k s_k(x')\right)}.$$

The probability is given for an empirically observed network x. Functions s_k are typically count statistics that evaluate the number of, for example, ties, reciprocal or homophilous ties, transitive structures, or in-degrees. They are thus similar to structural positions of network mechanisms after the inclusion of the explained tie. Coefficients β are associated with these structures and are subject to statistical estimation. If a graph has, for instance, a higher number of reciprocal structures than explained by the other parameters in the model (for example, the level of reciprocity is affected by the general density and the level of transitivity), the statistic will be associated with a positive coefficient β. More precisely, the parameters β can be interpreted as log-odds ratios. Positive (negative) values of a parameter indicate an over- (or under-) representation of a certain sub-graph structure, given the rest of the model, and therefore evidence for (or against) a certain network mechanism.

One challenge of estimating ERGMs is the denominator of the probability that sums the linear combination of statistics and parameters over the set $\mathcal{X}$ of all possible networks with n actors. A binary, directed network has $2^{n(n-1)}$ such outcomes. This number gets very large quickly when n is increased, which means that even for smaller networks with 10 or fewer nodes the probability cannot be calculated exactly. But iterative, approximate procedures exist that use Monte Carlo algorithms to estimate coefficients from data. Those routines are implemented in the "ergm" package in the R statistical environment (Hunter et al. 2008) and the program PNet (Wang et al. 2009), both freely available.

Table 23.3 summarizes empirical results, again based on the advice network of Lazega's lawyer study. Two models were estimated that are specified with network mechanisms as discussed in section 1.1.

Model 1 only includes network mechanisms that operate within a pair of actors. *Model 2* additionally includes network mechanisms that indicate that a tie variable may affect another tie variable even if those ties are not observed between the same two actors (as in the case of reciprocity).

We can interpret the parameters as follows. The *Density* parameter is a model intercept. It expresses the tendency of ties to be present in the first place. The fact that it is negative merely indicates that the overall density is significantly lower than 50%. Formally, the density effect can be understood as a network mechanism according to our definition, but one for which the structural position is empty and thus does not include any information. We further find evidence that ties are more likely to be observed when they are *reciprocated*. More senior people ask less for advice (negative effect for *Activity Seniority*), but they are asked more (positive *Attraction Seniority*). Lawyers tend to ask others for advice who are on the same level of seniority (*Homophily Seniority*) and who work in the same office (*Homophily Office*). No evidence is found for homophily in terms of coming from the same graduate school (*Homophily School*). The second model shows that ties tend to be transitively clustered (*Transitivity*) and that the network exhibits some in-degree centralization (*Popularity*). No evidence is found for a residual out-degree centralization (*Activity*).

Christoph Stadfeld and Viviana Amati - 9781789906851

Table 23.3 ERGM models investigating the advice network, considering different network mechanisms

	Model 1			Model 2		
Mechanism	Est.	s.e.	Sig.	Est.	s.e.	Sig.
Density	–3.81	0.12	***	–5.98	0.28	***
Reciprocity	1.17	0.14	***	0.99	0.14	***
Activity *Seniority*	–0.54	0.10	***	–0.63	0.10	***
Attraction *Seniority*	1.41	0.10	***	1.37	0.10	***
Homophily *Seniority*	0.92	0.09	***	0.76	0.07	***
Homophily *Office*	1.34	0.09	***	1.04	0.08	***
Homophily *School*	0.12	0.08		0.11	0.08	
Transitivity				1.66	0.20	***
Popularity				1.76	0.90	*
Activity				–0.79	0.97	

Notes: *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001.

A comparison of the two models illustrates why multi-mechanistic network models are useful beyond single-mechanistic tests that we illustrated in the section on CUG tests. We find that in Model 2, after the inclusion of transitivity, the estimates for homophily are smaller in each case. Transitivity in social networks can be explained in multiple ways, and one reason is indeed that actors are more likely to cluster with others who are similar (Granovetter 1973). Vice versa, homophily may be amplified by transitive clustering when, for example, actors are more likely to meet similar others due to the homophilous preferences of their friends. Once both mechanisms are represented in the model, the effects can be interpreted as the marginal effect, controlling for alternative explanations.

We conclude this section by mentioning that several extensions of ERGMs are available for two-mode networks, multivariate networks, and panel network data. Lusher et al. (2013) provide a comprehensive overview.

2.4 Stochastic Actor-Oriented Models

Stochastic actor-oriented models (SAOMs) are models for network panel data, i.e. observations of a network at M discrete time points $t_1, \ldots, t_M$. Conditioning on the first observed network x_{t_1}, they model a step-wise process explaining observed changes in the periods between the subsequent network data collections. The additional information captured by the timing of observations is exploited to test hypotheses on the network mechanisms driving the evolution of a network (Snijders 1996; 2001), the co-evolution between multiple networks (Snijders et al. 2013) or networks and actor attributes, referred to as behaviors (Steglich et al. 2010; Niezink et al. 2019).

Following the approach first proposed by Holland and Leinhardt (1977), SAOMs assume that the observed networks are discrete observations from an unobserved continuous-time

Christoph Stadfeld and Viviana Amati - 9781789906851

Markov process. The Markov assumption implies that the next configuration of the network depends only on the current configuration. This is a simplifying assumption that allows the decomposition of the network evolution into its simplest units, named mini-steps.

At each mini-step, a node in the network has the opportunity to revise one outgoing tie. Given the current state of the network x, the actor can decide to create a non-existing tie, dissolve an existing tie, or maintain the network unchanged.

The opportunity for a change is modeled by the rate function $\lambda_i(\alpha; x, v, w)$ describing the frequency per unit of time of change opportunities for an actor i. The rate might be constant over the actors (and be denoted by the parameter α), or depend on the position and attributes of the actors.

The choice of the selected actor i is modeled by the evaluation function $f_i(\beta; x, v, w)$ expressing the preference of the actor for each choice. The probability that the next state of the network is x' defined by a conditional multinomial model:

$$p_{(x')} = \frac{\exp\left(f_i(\beta; x', v, w)\right)}{\sum_{x''} \exp\left(f_i(\beta; x'', v, w)\right)}, \tag{23.1}$$

where the sum in the denominator is over all possible options in the mini-step.

The evaluation function is defined as a linear combination $f_i(\beta; x', v, w) = \Sigma_k \beta_k s_{ki}(x', v, w)$ of statistics $s_{ki}(x', v, w)$ and statistical parameters β_k which can be interpreted as log-probability ratios. The statistics s_{ki} are counts of local configurations similar to the ones in ERGMs but computed from the point of view of actor i (Block et al. 2019). They can thus be viewed as definitions of structural positions of a network mechanism. The parameter β_k expresses the preference of actor i for ties leading to the local configuration of type k. Extensions of SAOMs to the co-evolution of networks and behaviors, and multiple networks are obtained by defining an evaluation function for each dependent variable and introducing statistics s_{ik} that depend simultaneously on multiple networks or networks and behaviors.

Snijders (1996) justified the formula in equation (23.1), building on concepts of random utility theory. Actors can be considered as decision-makers. When they get the opportunity to make a tie change, they choose the option of maximizing their utility. The utility is defined as the sum of the evaluation function, which represents the payoff of the nodes when making a particular change, and the error term, whose distributional assumption determines the probabilities in Equation (23.1). It should be noted that similar to the use of multinomial regression models in other contexts, the interpretation of parameters does have to assume a strict utility maximization, but it can merely focus on the fact that tie creations of an actor in certain structural positions are more likely than in others. Yet, it is important to review behavioral assumptions imposed by a specific model and assess whether they are reasonable in a given empirical context.

The quantity α and β are statistical parameters that need to be estimated from the data. Several estimation procedures have been developed and implemented in the R package "RSiena" (Ripley et al. 2020). Since the exact sequence of opportunities for changes is not observed, analytical solutions cannot be computed. A stochastic algorithm is used instead to compute the parameter estimates (Snijders 2001).

We use the data from the Swiss StudentLife study to illustrate SAOMs. We consider a model that analyses the co-evolution of the friendship (on the left) and the informational support (on

Christoph Stadfeld and Viviana Amati - 9781789906851

Table 23.4 SAOM investigating the co-evolution between friendship and informational support. The models are coupled and consider network mechanisms, attribute-related mechanisms, and whether one network explains ties in the other

	Friendship			Support		
Mechanism	Est.	s.e.	Sig.	Est.	s.e.	Sig.
Rate function						
Rate	11.05	0.97		9.11	1.12	
Evaluation function						
Density	–3.28	0.38	***	–4.34	0.50	***
Reciprocity	3.05	0.30	***	2.20	0.35	***
Transitivity	0.87	0.08	***	0.85	0.22	***
Transitivity x Reciprocity	–0.81	0.13	***	–0.78	0.36	*
Popularity	0.35	0.12	**	0.46	0.15	**
Out-Popularity	–0.65	0.12	***	–0.78	0.19	***
Activity	0.01	0.08		0.10	0.13	
Activity *Female*	0.16	0.11		0.17	0.16	
Attraction *Female*	0.37	0.10	***	0.25	0.13	*
Homophily *Gender*	0.11	0.12		0.18	0.15	
Homophily *Study program*	0.71	0.12	***	0.84	0.14	***
Other network	2.37	0.50	***	2.12	0.36	***

Notes: *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001.

the right) networks. The model is estimated from two survey waves of network data collected among a cohort of undergraduate students at a Swiss university. Data were collected at the beginning and after one month of their studies. The model conditions on the friendship and support networks of the first wave thus explain the change observed in the second survey wave (i.e. the ties being maintained, added and dropped). Table 23.4 reports the estimated coefficients.

The *Rate* parameter describes the sub-process that governs the overall "speed" of the process. Estimates of 11 and 9 indicate that on average each student considered changing their friendship and support ties 11 and 9 times, respectively.

The parameters of the evaluation function are interpreted as actors' preferences. For the friendship networks, the parameters related to *Density* and *Reciprocity* reflect a preference of students to create or maintain a low number of ties and ties that reciprocate existing ties. The positive *Transitivity* parameter suggests that students tend to befriend their friends' friends. However, the negative parameter of the interaction between transitivity and reciprocity (*Transitivity* × *Reciprocity*) indicates that the 'utility' of transitivity and reciprocity does not add up linearly, such that reciprocal ties within triads are not much preferred over non-reciprocated, transitive ties.

Christoph Stadfeld and Viviana Amati - 9781789906851

The other significant parameters indicate that students have preferences to befriend popular students (*Popularity*), female students (*Attraction Female*), and those enrolled in the same study program (*Homophily Study program*). Similar interpretations holds for the evolution of the support network.

The last effect in the table (*Other network*) expresses whether the evolutionary processes of friendship and support networks are intertwined, i.e. friendship ties are explained by support ties and vice versa. The positive and significant coefficients of the other network effects suggest that students who maintain a tie in one network (friendship or support) are more likely to create or maintain a tie in the other network as well. The SAOM is a multi-mechanistic network model that tests network mechanisms using temporal information that is typically not used in ERGM analyses.

2.5 Relational Event Models

Sometimes network data are collected in the form of relational events, i.e. time-stamped sequences of dyadic observations. Such data stem from a variety of sources, like archival network data on treaties, trades and interactions, social sensor experiments, digital communication, and social media platforms. Formally, relational event data are sequences of time-ordered triplets: $w_k = (t_k, i_k, j_k)$, where t_k is the time, i_k the sender, and j_k the receiver of the event. The three entries can indicate, for example, which two parties signed a treaty at a certain date (an undirected event), or who called whom at a certain point in time (a directed event).

Relational event models build on statistical network models like the ERGMs and SAOMs. They were first proposed by Butts (2008), as models of tie hazards that are similarly specified to tie-oriented ERGMs, and later reformulated in the actor-oriented tradition of SAOMs (e.g. Stadtfeld and Geyer-Schulz 2011; Vu et al. 2011), deconstructing the model into sender activity rates and multinomial probabilities to express the choice of event receivers by event senders.

Both approaches have in common that they consider not only the order of events but their actual timing by formulating tie hazards λ_{ij}. The higher the hazard between two actors, the shorter the expected time span between two subsequently observed events. The hazards may vary due to prior observed events (inertia), attributes of sender and receiver (e.g. homophily), the presence of ties in another network (e.g. perceived closeness), or the relative network positions of sender and receiver (e.g. reciprocity, transitivity, popularity).

Hazards for tie- and actor-oriented relational event models can be formulated as follows:

$$\lambda_{ij}^{\text{Tie}}(x,\gamma,u,A) = \exp(\gamma_0 + \gamma^T u(x,i,j)) \tag{23.2}$$

$$\lambda_{ij}^{\text{Actor}}(x,\theta,\beta,s,t,A) = \underbrace{\exp(\theta_0 + \theta^T s(x,i))}_{\text{step 1: sender rate}} \underbrace{\frac{\exp(\beta^T t(x,i,j))}{\sum_{k\in A}\exp(\beta^T t(x,i,k))}}_{\text{step 2: receiver choice}}. \tag{23.3}$$

Variables i, j indicate the sender and receiver of the event; x the process state at a given time, including information about the history of the process, attributes, other networks, and global variables; γ, θ, β are coefficients to be estimated from the data; u, s, t are functions evaluating

the position of the events in the process state that are used to measure the structural positions of network mechanisms; *A* is the set of nodes between which events are considered. Details on actor-oriented relational event models (dynamic network actor models, DyNAMs) are provided in Stadtfeld et al. (2017).

We illustrate the application of a DyNAM model using the Swiss StudentLife study data. An introductory event for new undergraduate students took place at the end of the first week of their studies. The goals of this event were for students to mingle and meet others, to talk to senior students who could give advice, and to attend presentations by professors and administrators. Students who participated in the event were from the same department and year group as the student sample used in the SAOM section. However, they followed a different major. Participants were wearing RFID badges that recorded the exact time of face-to-face proximity (for details, see Elmer et al. 2019). Overall, more than 12 000 events were recorded between 50 participating individuals over about 21 hours. The events are undirected. However, for illustrative purposes, we treat them as directed after randomizing their direction (although models for undirected relational event data are available, see Stadtfeld et al. (2017)). The DyNAM model answers two general questions: which students were most likely to interact and with whom?

Table 23.5 shows the results of the DyNAM model. The upper part includes coefficients of the *rate* function, the lower part of the table the *choice* function.

The rate model is concerned with structural positions that describe the sender of the event. The intercept expresses the expected waiting time between two events of the same individuals assuming all other statistics are zero. Individual activity is partly explained by recent activity within the last 10 minutes and by having more ties in the friendship network. Beyond these effects, we find no evidence that individuals with many interaction ties in the past are more likely to interact in the future (the marginal Activity effect is negative).

The choice part of the model is concerned with structural positions that evaluate the person being chosen as an interaction partner and the relative position and similarity of the sender and receiver. Individuals tend to interact with the same people at the event both in general (*Inertia*) and in the short term (*Inertia 10 Minutes*). The same pattern holds for transitivity. Individuals tend to interact with former interaction partners of their former interaction partners both in general and in the short term (*Transitivity*, *Transitivity 10 Minutes*). They further tend to interact with friends of their friends (*Transitivity Friendship*) on top and beyond their tendency to interact with their friends (*With Friends*). Popularity effects resemble the Activity effects in the rate function: individuals tend to interact with those who interact a lot in the short term (*Popularity 10 Minutes*), but not necessarily with those who have a high interaction degree over the whole event (*Popularity*). Again, this may indicate that "preferential attachment" is naturally limited in the context of a student welcome event where most people meet many others. There is no gender attractivity effect, but we find evidence for the opposite mechanism of homophily: interaction events are more likely to occur between students of the opposite gender (negative *Homophily Gender* effect).

The model aims at highlighting how structural positions of network mechanisms can be refined when specific detailed data are available. Here, we tested how network mechanisms operate differently in the short-term and long-term (see also Amati et al. 2019). Other refinements of network mechanisms are similarly possible, for example by considering situational measurements on the network level (e.g. time of the day), or individual level (e.g. stress) (Stadtfeld et al. 2017).

Christoph Stadfeld and Viviana Amati - 9781789906851

Table 23.5 DyNAM model investigating the sender rates and receiver choices in face-to-face interactions. The model controls for network positions in the interaction and friendship networks, long-term and short-term (10 minutes) interactions, and gender

Mechanism	Est.	s.e.	Sig.
Rate			
Intercept	–5.99	0.03	
Activity	–0.01	0.00	***
Activity *10 Minutes*	0.25	0.00	***
Activity *Friendship*	0.03	0.01	***
Activity *Female*	0.00	0.02	
Choice			
Inertia	1.09	0.04	***
Inertia *10 Minutes*	0.81	0.03	***
Transitivity	0.05	0.01	***
Transitivity *10 Minutes*	0.47	0.01	***
Transitivity *Friendship*	0.07	0.02	***
Popularity	–0.02	0.00	***
Popularity *10 Minutes*	0.05	0.01	***
Popularity *Friendship*	–0.04	0.01	***
With friends	0.91	0.04	***
Attractivity *Female*	–0.02	0.02	
Homophily *Gender*	–0.05	0.02	*

Notes: * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.

3. NETWORK MODELS FOR SIMULATION

The presented network models partly provide multi-mechanistic explanations of tie observations in social networks. The explanations are on the micro level and consider the positions of individuals in a potentially complex social network. In this section, we review how agent-based simulations can be conducted from these micro-level models to investigate the micro-macro link.

Several of the discussed statistical network model types can be interpreted as the outcome of a generative process. Partly, this process is explicit in the formulation of the model (SAOM, REM, DyNAM); partly it can be implicitly derived (ERGM). This enables us to apply these statistical network models as simulation models or *agent-based models* (ABMs, see Manzo 2014; Bianchi and Squazzoni 2015).

While statistical network models are explicit about the social network mechanisms operating on the individual level, it is often unclear how those micro-level mechanisms can be related to outcomes observed on the network macro-level. Simulation studies using information from

Christoph Stadfeld and Viviana Amati - 9781789906851

empirical network models can help to establish this micro-macro link in social networks (Snijders and Steglich 2015). The question of how micro mechanisms relate to macro-level features of social systems and vice versa is at the core of analytical sociology. The formalized language that social networks and social network models provide can help to study them in a rigorous manner (Stadtfeld 2018).

We illustrate the utility of this approach by using the SAOM models estimated as simulation models. The simulations start from the final observed empirical network and follow the logic of the SAOM formulation: actors are randomly chosen to reconsider one of their current network ties. They can then add a tie to someone they are not yet connected to, they can drop one of their existing ties, or they can keep the network unchanged. This choice is expressed by the probability in Equation 23.1 The coefficients β are taken from empirically estimated models. The length of the simulation is governed by the rate parameters, which represent the expected numbers of change considerations per actor. To induce more variability in the process, we simulate four times the number of changes compared to the estimated models. As data were estimated from one month of data, the simulated data thus roughly represent a four-month period.

We address the question of how the estimated *homophily* network mechanisms affect features of the emerging social network. While the current example only considers one specific network, we may define similar simulation studies that for example consider the co-evolution of multiple networks and individual attributes, thereby representing complex macro-level outcomes of social systems. We compare two ABM models: one is the estimated model in Table 23.4 (full model), one model is identical to the first except for the fact that no homophily mechanism was included in the estimation (reduced model).

Figure 23.3 shows exemplary simulated networks. One apparent difference between the networks generated by the two models is that in the full model clustering is very much aligned with attributes (study major). This is not the case for networks generated from the reduced model where homophily of study majors was not considered. From visual inspection, no differences in network clustering appear.

Below the network plots, network features are compared more systematically, based on 200 simulations per model. Statistics are calculated for the joint friendship and support network. It can be seen that the network density is slightly higher in the full model (2.5% vs. 2.2%) and that, unsurprisingly, about twice as many homophilous ties exist in the networks generated from the full models (83% vs. 40%). The modularity (a feature that expresses the tendency of network ties to be observed within densely connected subgroups) of networks generated under the full model is higher on average (0.50 vs. 0.46). Finally, the full model generates a higher number of fully connected triads (cliques of size 3) in which among three people everybody is connected to everybody else reciprocally (252 vs. 161). The simulations thus illustrate how network-level properties are affected by the inclusion of network mechanisms on the individual level. One may consider that emerging outcomes, such as the modularity of the network of the similarity of connected nodes, can affect and constrain how individuals will act and interact in the future.

Similarly, we could use agent-based simulations of statistical network models to conduct "counter-factual" thought experiments, such as considering that certain mechanisms were stronger or not present at all, that the observation period was longer, or that the initial conditions from which the network evolved were different (e.g. considering different initial levels of connectedness or distributions of attributes). The literature on agent-based models can provide

Christoph Stadfeld and Viviana Amati - 9781789906851

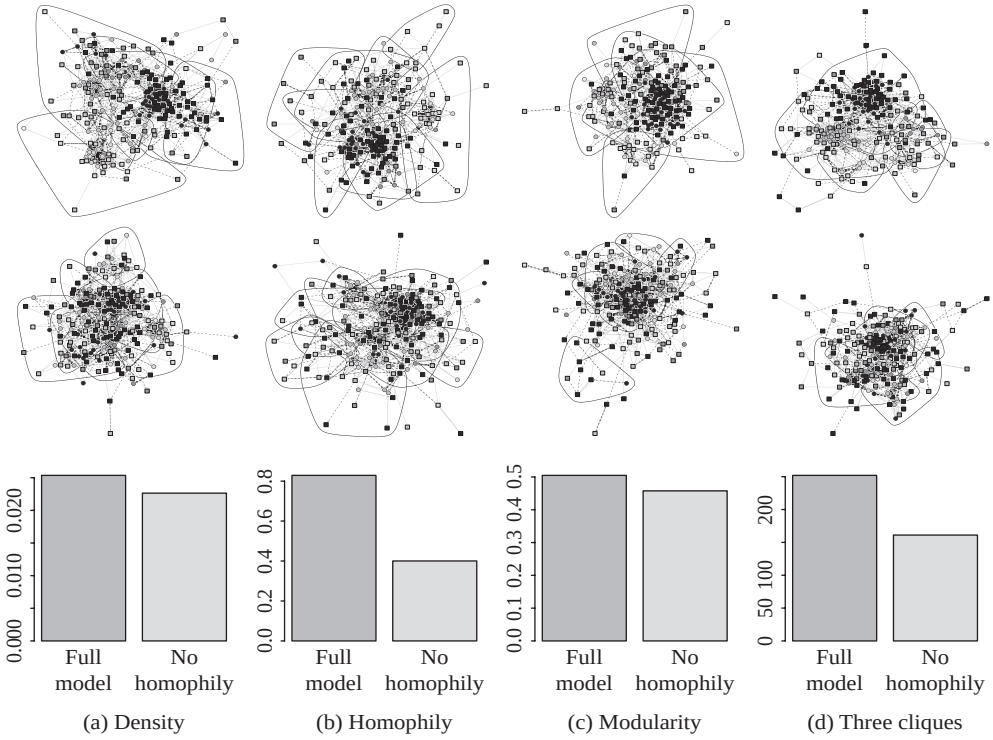


Figure 23.3 First row: Example networks simulated from the full SAOM model; second row: Example networks simulated from the reduced SAOM model re-estimated without the homophily mechanisms; colors and shapes as described in Figure 23.1, circles in the background indicate detected communities based on the modularity criterion; third row: Comparison of macro-level network statistics (full model: dark gray; reduced model: light gray)

further inspiration for the design of such empirically informed simulation studies that investigate emerging macro-level phenomena. The empirical description of social phenomena and the formalization and investigation of generative processes through simulation experiments are indeed two important aspects of the analytical path proposed by Manzo (2014). Social network concepts and multi-mechanistic network models are a promising way to integrate both into a single analytical framework.

4. DISCUSSION

This chapter defined *network mechanisms*—widely observed classes of structural processes in social networks—and compared them to social mechanisms as introduced by Hedström and Swedberg (1998), and further formalized by Hedström and Ylikoski (2014), Manzo (2014), and others. It then summarized and explained different statistical *network models*. These can be applied to test whether evidence for network mechanisms can be found in empirical social network data and, in particular, to investigate their residual effect when multiple network

mechanisms are incorporated into one model. By doing this, researchers can gain a deeper understanding of the processes that explain the structure and dynamics of social networks in a variety of empirical settings, ranging from small-scale survey studies to large-scale studies using digital data formats. One strength of many modern network models is that they can both be applied as explanatory statistical models and as agent-based simulation models. Empirical examples illustrated how the application of network models may open new insights into the micro-level processes in social networks and their role in emerging macro-level phenomena.

REFERENCES

Amati, Viviana, Alessandro Lomi and Daniele Mascia (2019), "Some days are better than others: Examining time-specific variation in the structuring of interorganizational relations", *Social Networks*, **57**: 18–33.

Barabási, Albert-László and Réka Albert (1999), "Emergence of scaling in random networks", *Science*, **286** (5439): 509–12.

Bianchi, Federico and Flaminio Squazzoni (2015), "Agent-based models in sociology", *Wiley Interdisciplinary Reviews: Computational Statistics*, **7** (4): 284–306.

Block, Per, Christoph Stadtfeld and Tom A.B. Snijders (2019), "Forms of dependence: Comparing SAOMs and ERGMs from basic principles", *Sociological Methods and Research*, **48** (1): 202–39.

Brandes, Ulrik (2016), "Network positions", *Methodological Innovations*, **9**: 1–19.

Brandes, Ulrik, Garry Robins, Ann McCranie and Stanley Wasserman (2013), "What is network science?", *Network science*, **1** (1): 1–15.

Butts, Carter T. (2008), "A relational event framework for social action", *Sociological Methodology*, **38** (1): 155–200.

Coleman, James Samuel (1990), *Foundations of Social Theory*, Cambridge, MA: Belknap Press.

Dekker, David, David Krackhardt and Tom A.B. Snijders (2007), "Sensitivity of MRQAP tests to collinearity and autocorrelation conditions", *Psychometrika*, **72** (4): 563–81.

Elmer, Timon, Krishna Chaitanya, Prateek Purwar and Christoph Stadtfeld (2019), "The validity of RFID badges measuring face-to-face interactions", *Behavior Research Methods*, **51** (5): 2120–38.

Emerson, Richard M. (1976), "Social exchange theory", *Annual Review of Sociology*, **2** (1): 335–62.

Erdős, P. and A. Rényi (1959), "On random graphs I", *Publicationes Mathematicae*, **6** (1): 290–7.

Gilbert, Edgar N. (1959), "Random graphs", *The Annals of Mathematical Statistics*, **30** (4): 1141–4.

Granovetter, Mark S. (1973), "The strength of weak ties", *American Journal of Sociology*, **78** (6): 1360–80.

Hedström, Peter and Peter Bearman (2009), *What is Analytical Sociology All About? An Introductory Essay*, in Peter Hedström and Peter Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press, pp. 3–24.

Hedström, Peter and Richard Swedberg (1998), "Social Mechanisms: An Introductory Essay", in Peter Hedström and Richard Swedberg (eds), *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge: Cambridge University Press, pp. 1–31.

Hedström, Peter and Petri Ylikoski (2014), "Analytical Sociology and Rational-Choice Theory", in Gianluca Manzo (ed.), *Analytical Sociology*: *Actions and Networks*, New York: Wiley, pp. 57–70.

Heider, Fritz (1958), *The Psychology of Interpersonal Relation*, New York: Wiley.

Hoff, Peter D., Adrian E. Raftery and Mark S. Handcock (2002), "Latent space approaches to social network analysis", *Journal of the American Statistical Association*, **97** (460): 1090–8.

Holland, Paul W., Kathryn Blackmond Laskey and Samuel Leinhardt (1983), "Stochastic blockmodels: First steps", *Social Networks*, **5** (2): 109–37.

Holland, Paul W. and Samuel Leinhardt (1977), "A dynamic model for social networks", *Journal of Mathematical Sociology*, **5** (1): 5–20.

Hunter, David R., Mark S. Handcock, Carter T. Butts, Steven M. Goodreau and Martina Morris (2008), "ergm: A package to fit, simulate and diagnose exponential-family models for networks", *Journal of Statistical Software*, **24** (3): 1–29.

Huston, Ted L. and George Levinger (1978), "Interpersonal attraction and relationships", *Annual Review of Psychology*, **29** (1): 115–56.

Krackhardt, David (1988), "Predicting with networks: Nonparametric multiple regression analysis of dyadic data", *Social Networks*, **10** (4): 359–81.

Lazarsfeld, Paul F. and Robert K. Merton (1954), "Friendship as a social process: A substantive and methodological analysis", in Morroe Berger, Theodore Abel, and Charles H. Page (eds), *Freedom and Control in Modern Society*, New York: van Nonstrand, pp. 18–66.

Lazega, Emmanuel (2001), *The Collegial Phenomenon: The Social Mechanisms of Cooperation among Peers in a Corporate Law Partnership*, Oxford University Press on Demand.

Christoph Stadfeld and Viviana Amati - 9781789906851

Lusher, Dean, Johan Koskinen and Garry Robins (2013), *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*, Cambridge: Cambridge University Press.
Manzo, Gianluca (2007), "Variables, mechanisms, and simulations: Can the three methods be synthesized?", *Revue française de sociologie*, **48** (5): 35–71.
Manzo, Gianluca (2014), "Data, Generative Models, and Mechanisms: More on the Principles of Analytical Sociology", in Gianluca Manzo (ed.), *Analytical Sociology: Actions and Networks*, New York: Wiley, pp. 4–52.
McPherson, J. Miller, Lynn Smith-Lovin and John Manuel Cook (2001), "Birds of a feather: Homophily in social networks", *Annual Review of Sociology*, **27**: 415–44.
Merton, Robert K. (1968), "The Matthew effect in science", *Science*, **159** (3810): 56–63.
Niezink, Nynke M.D., Tom A.B. Snijders and Marijtje A.J. van Duijn (2019), "No longer discrete: Modeling the dynamics of social networks and continuous behavior", *Sociological Methodology*, **49** (1): 295–340.
Ripley, Ruth M., Tom A.B. Snijders, Zsófia Boda, András Vörös and Paulina Preciado (2020), *Manual for Siena version 4.0*. Tech. rep., Oxford: University of Oxford, Department of Statistics; Nuffield College. https://www.cran.r-project.org/web/packages/RSiena/.
Robins, Garry, Pip Pattison, Yuval Kalish and Dean Lusher (2007) "An introduction to exponential random graph (p*) models for social networks", *Social Networks*, **29** (2): 173–91.
Roux, Valentine and Gianluca Manzo (2018), "Social boundaries and networks in the diffusion of innovations: A short introduction", *Journal of Archaeological Method and Theory*, **25** (4): 967–73.
Snijders, Tom A.B. (1996), "Stochastic actor-oriented models for network change", *Journal of Mathematical Sociology*, **21** (1–2): 149–72.
Snijders, Tom A.B. (2001), "The statistical evaluation of social network dynamics", *Sociological Methodology*, **31** (1): 361–95.
Snijders, Tom A.B., Alessandro Lomi and Vanina Jasmine Torló (2013), "A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice", *Social Networks*, **35** (2): 265–76.
Snijders, Tom A.B., Gerhard G. Van de Bunt and Christian E.G. Steglich (2010), "Introduction to stochastic actor-based models for network dynamics", *Social Networks*, **32** (1): 44–60.
Snijders, Tom A.B. and Christian E.G. Steglich (2015), "Representing micro-macro linkages by actor-based dynamic network models", *Sociological Methods and Research*, **44** (2): 222–71.
Stadtfeld, Christoph (2018), "The Micro-Macro Link in Social Networks", in Robert A. Scott and Marlis C. Buchmann (eds), *Emerging Trends in the Social and Behavioral Sciences*, New York: Wiley.
Stadtfeld, Christoph and Andreas Geyer-Schulz (2011), "Analyzing event stream dynamics in two-mode networks: An exploratory analysis of private communication in a question and answer community", *Social Networks*, **33** (4): 258–72.
Stadtfeld, Christoph, James Hollway and Per Block (2017), "Dynamic network actor models: Investigating coordination ties through time", *Sociological Methodology*, **47** (1): 1–40.
Stadtfeld, Christoph, András Vörös, Timon Elmer, Zsófia Boda and Isabel J. Raabe (2019), "Integration in emerging social networks explains academic failure and success", *Proceedings of the National Academy of Sciences* (PNAS), **116** (3): 792–7.
Steglich, Christian, Tom A.B. Snijders and Michael Pearson (2010), "Dynamic networks and behavior: Separating selection from influence", *Sociological Methodology*, **40** (1): 329–93.
Vörös, András, Zsófia Boda, Timon Elmer, Marion Hoffman, Kieran Mepham, Isabel J. Raabe and Christoph Stadtfeld (2021), "The Swiss StudentLife study: Investigating the emergence of an undergraduate community through dynamic, multidimensional social network data", *Social Networks*, **65**: 71–84.
Vu, Duy Q., David Hunter, Padhraic Smyth and Arthur U. Asuncion (2011), "Continuous-time regression models for longitudinal networks", in *Advances in Neural Information Processing Systems*, pp. 2492–500.
Wang, Peng, Garry Robins and Philippa Pattison (2009), *Pnet: Program for the Simulation and Estimation of Exponential Random Graph (p*) Models: User Manual*, Melbourne, Australia: University of Melbourne.
Wasserman, Stanley and Katherine Faust (1994), *Social Network Analysis: Methods and Applications*, Vol. 8, Cambridge: Cambridge University Press.
Watts, Duncan J. (2004), "The 'new' science of networks", *Annu. Rev. Sociol*, **30**: 243–70.
Watts, Duncan J. and Steven H. Strogatz (1998), "Collective dynamics of 'small-world' networks", *Nature*, **393** (6684): 440–2.

24. Agent-based computational models[1]

Andreas Flache and Carlos A. de Matos Fernandes[2]

1. INTRODUCTION

Agent-based computational modeling (ABCM) plays a central role in analytical sociology (Hedström and Bearman, 2009). Analytical sociology (AS) aims to explain social phenomena with precisely described and empirically realistic social mechanisms. Mechanistic explanations in AS target, in particular, complex micro-macro interactions in which individual behavior at the micro level of society can, in interaction with meso- and macro-level structures, bring about macro outcomes individuals neither necessarily anticipate nor desire. Examples are residential segregation (Clark and Fossett, 2008), opinion polarization (Flache, Mäs, et al., 2017), or dysfunctional status hierarchies (Mark et al., 2009). AS aspires to uncover why this occurs in real-life social settings by modeling the underlying social mechanism at a level of precision that allows logical deduction and empirical refutation of both model assumptions and predictions. With this ambition, AS is not fundamentally different from, for example, empirically oriented Rational Choice Sociology. However, despite the considerable flexibility of modern conceptualizations of Rational Choice Sociology (Wittek et al., 2013), AS is different in that is does not commit to a particular set of ingredients for modeling individual behavior (Manzo, 2014, pp. 21–27). ABCM's capacity to combine high theoretical flexibility with analytical precision is exactly what makes it so attractive to analytical sociologists.

Computer simulation is the main approach in ABCM for analyzing whether, and if so, how, and under which conditions exactly the social mechanism formulated in a computational model can generate the phenomenon the researcher wants to understand (Epstein, 2006). While ABCM is not the only method for formalizing social mechanisms, it is particularly compatible with AS (Macy and Flache, 2009; Manzo, 2014, 2021, *forthcoming*) because it combines formal precision and the ability to model and analyze complex micro-macro interactions (Mäs, Chapter 4, this volume) and theoretical flexibility in a unique way.

ABCM forces a modeler to make the explanatory mechanism fully explicit. The need to translate a conceptual mechanistic model into an executable computer program requires describing algorithmically how agents arrive at changes in their actions, cognitions or emotions

[1] This is an open-access chapter distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 Unported (https://creativecommons.org/licenses/by-nc-nd/4.0/). Users can redistribute the work for non-commercial purposes, as long as it is passed along unchanged and in whole, as detailed in the License. Edward Elgar Publishing Ltd. must be clearly credited as the rights holder for publication of the original work. Any translation or adaptation of the original content requires the written authorization of Edward Elgar Publishing Ltd.

[2] The first author gratefully acknowledges financial support by the Netherlands Organization for Scientific Research (NWO) under the 2018 ORA grant ToRealSim (464.18.112). Both authors acknowledge that this study is part of the research program Sustainable Cooperation—Roadmaps to Resilient Societies (SCOOP) funded by NWO and the Dutch Ministry of Education, Culture and Science (OCW) in its 2017 Gravitation Program (grant number 024.003.025). Further, we are grateful to Gianluca Manzo for thoughtful feedback and editorial support, as well as to the members of the Norms and Networks Cluster (NNC) at the University of Groningen, Department of Sociology, for an inspiring and helpful discussion of an earlier version of this chapter.

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

interdependently with their local and global social context. This is different from the approach taken by game theory, where individual actions are seen as part of a strategy profile in a game that satisfies axiomatic requirements for strategically rational choice, while the exact cognitive process that leads actors to make those choices is left implicit (Breen, 2009). Formal approaches with less demanding cognitive assumptions like evolutionary game theory (Bendor and Swistak, 2001), Markov chain models (Young, 1998) or socio-physics (Castellano et al., 2009) are similarly capable of addressing complex micro-macro interactions and offer in addition powerful analytical tools that can yield more general insights into model implications than ABCM. However, these approaches may require a modeler to press their substantive theory of a social mechanism into the Procrustean bed of a particular formal framework. From the AS perspective this is undesirable because realistic models of social mechanisms should only be constrained by the theoretical and empirical information at hand, as well as the requirement of logical consistency and precision. Keuschnigg et al. (2018, p. 6) express this view as follows:

> It is not enough that the theory or model predicts the phenomena to be explained; an explanation must represent the essential features of the actual process that produced the phenomena to be explained. […] from the AS perspective, considerations of elegance, simplicity, or tractability never should override the aim of accurately describing the mechanisms that actually produced the phenomena to be explained.

Analytic precision, ability to capture complex micro-macro interactions, and flexibility to accommodate empirically realistic assumptions suggest for many adherents of AS that ABCM is a highly suitable approach to provide mechanistic explanations. But this does not mean that anything goes. Other than analytical approaches like game theory or socio physics, ABCM typically does not come with the possibility to mathematically prove that model results follow correctly from model assumptions. Empirically plausible models that mimic real phenomena are no guarantee that the process by which the model generated the result has been sufficiently understood by the modeler to exclude artefacts created by inessential model details or perhaps programming errors. ABCM is most needed when the implications of a seemingly simple mechanism are too complex for a researcher to be sufficiently understood without formal modeling. But this also implies that a model built to capture a complex social mechanism can be a complex system in itself that needs to be understood well in the first place. Without the full comprehension of the dynamic processes generated by an ABCM, little has been gained if these dynamic processes resemble the social phenomena the researcher wants to explain. In other words, explaining the outcomes generated by the ABCM itself is different from but logically precedes explaining a social phenomenon through that ABCM.[3] Yet, there is still a major gain from creating a complex ABCM for explaining a complex phenomenon. Other than social reality, the computational model is fully transparent and fully under the control of the researcher, allowing for systematic analysis and experimentation with it. Thus, the potential of ABCM to aid AS can be fully realized if combined with adherence to a set of good practices for systematic deductive model development and analysis.

This chapter aims to give readers an easily accessible introduction to ABCM in the context of AS. Gaining practical experience is crucial for understanding the benefits and potential

[3] We thank Gianluca Manzo for having inspired this formulation.

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

pitfalls of ABCM. A core element of the chapter is, therefore, the development of a modeling example the reader can replicate themself with *NetLogo* (Wilensky, 1999), a freely available software. We start from Schelling's famous model of segregation, the same "toy-model" (Reutlinger et al., 2017) that Hedström and Bearman (2009) used to illustrate how ABCM helps to understand complex micro-macro interactions.[4] We extend this model toward including more empirically informed assumptions. This demonstrates the use of the good practices we propose for ABCM, showing how it can be challenging in ABCM to understand what the consequences of adding realism to a model are, and how navigating between simpler, less realistic models, and more complex models with a higher degree of realism is needed for full comprehension of model outcomes.

In what follows, we first describe the main ingredients of an ABCM (section 2). Section 3 presents the modeling example and good practices. We conclude the paper with a broader reflection of the role of ABCM in AS.

2. A BIRD'S EYE VIEW ON ABCM

What follows is a brief look at the main components of an ABCM. Comprehensive overviews and discussions of ABCM in sociology, and AS in particular, can be found in the literature (Macy and Flache, 2009; Squazzoni, 2012; Manzo, 2014; Bianchi and Squazzoni, 2015).

The core element of an ABCM are **agents**, software objects representing individual social actors construed simultaneously as *autonomous*, *interdependent*, *heterogeneous* and *embedded* (Wooldridge and Jennings, 1995). Agents comprise state variables, representing fixed characteristics like their sex or skin color, but also beliefs, opinions, emotions, or actions that agents change autonomously, depending on their current state and on external inputs. From the perspective of AS, agent-level action variables must be construed such that the actions of all agents jointly constitute **macro-level outcomes**, mirroring the empirical social phenomenon of interest. An example is residential segregation, a macro-outcome computed from the distribution of individual agents over the space of an artificial city, which in turn is the result of agents' residential choices and moving actions (Benenson et al., 2009). The initial value of state variables, representing internal agent states, beliefs, opinions or initial residential locations, as well as initial macro-level characteristics like the size of an agent population, can be seen as **modeling the social situational context** of which the modeler wants to understand why and how it is related to the social outcome of interest.

A **model of agent behavior** in an ABCM comprises both assumptions about agents' internal states and actions as well as assumptions about their social interdependence. In modeling decision-making, ABC modelers typically impose comparatively low demands on the cognitive ability of agents, but explicate agent's decision-making process, motivated by critique of the descriptive inaccuracies of the axiomatic behavioral model of perfectly rational

[4] A slightly updated version of the model Hedström and Bearman (2009) used is available in the NetLogo model library under the name *segregation.nlogo* (Wilensky, 1997). The models presented in this chapter are considerably further developed to include empirically more realistic assumptions. Following an important good practice in ABCM, we make our code available for replication. We provide two *NetLogo* models of segregation, called *SegregationExtended* and *SegregationDiscreteChoice*, respectively, available for download at https://www.comses.net/codebase-release/db473a1d-c803-4e9d-8b2f-d9389784a063/. To use the models, first download and install the software NetLogo (https://ccl.northwestern.edu/netlogo/download.shtml, version 6.1.1.).

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

choice (Macy and Flache, 2009). Bounded rationality (Simon, 1982), heuristic decision-making (Goldstein, 2009), simple reinforcement learning (Macy and Flache, 2002) or more sophisticated forms of learning (Zschache, 2018) are common approaches used in modeling individual-level behavior in an ABCM.

While autonomous agents control their own internal states, their autonomy is constrained by **interdependence**. Each agent's actions and possibilities for future actions depend on what happens in the agent's local environment, and these actions in turn have consequences that alter their own environment and that of others. Important for many applications in the social sciences, agents are locally interdependent, for example in a social network that links them with a (typically small) subset of the overall population, such that macro-level outcomes are an emergent property of local interactions. Local interaction is often seen as a key ingredient to a realistic description of social dynamics in large populations (Axelrod, 1997). In addition, ABCMs can include assumptions about how agents change their position in a geographical space or break off relations and establish new ones with other agents (Centola et al., 2007; Bianchi et al., 2020).

ABCMs have been applied to study the emergence and dynamics of a wide range of social phenomena, including residential segregation (Schelling, 1971; Clark and Fossett, 2008; Hegselmann, 2017), cultural differentiation (Axelrod, 1997; Flache and Macy, 2011), norm diffusion (Mäs and Opp, 2016), opinion polarization (Baldassarri and Bearman, 2007; Flache et al., 2017) or the diffusion of innovations (Manzo et al., 2018). In the next section, we use a model of residential segregation to illustrate good practices and potential pitfalls generally applicable to ABCM.

3. AN ABCM OF UNINTENDED RESIDENTIAL SEGREGATION: FROM A TOY-MODEL TO A (SOMEWHAT MORE) REALISTIC MODEL

Schelling (1971) famously proposed an extremely simple but highly illustrative social mechanism to understand how strong ethnic segregation could arise in a world where individuals do not actually want it. Schelling's simple computational model will be taken as the starting point for our modeling example, but readers should be aware that independently different models implementing essentially the same mechanism have been proposed (Sakoda, 1971; Hegselmann, 2017). The core part of this mechanism is a *relocation cascade*. Even if individuals are content with living in a mixed neighborhood, they also want at least a certain minimal fraction of co-ethnics nearby. As soon as some individuals feel outnumbered, they leave and settle elsewhere. In the process, movers unintendedly make it more likely that other members of their group follow suit, because after they leave, their group becomes even more outnumbered in their prior neighborhood. Multiple modeling studies have shown how this simple micro-level process can induce a self-reinforcing cascade in which more and more relocations occur until neighborhoods end up being highly segregated (for an overview, see Hegselmann, 2017).

3.1 Social Context and Macro-Outcomes: Which Social Phenomenon Is to Be Explained?

AS wants to provide precise and realistic mechanistic explanations of concrete and puzzling empirical macro-phenomena. Schelling's model, in particular, addressed what he saw as a

puzzle. Ethnically diverse cities in the US (and elsewhere) are highly ethnically segregated, although empirical studies suggest that most people would prefer living in mixed neighborhoods (Clark and Fossett, 2008).

ABCM forces a researcher to formulate precise, formal descriptions of all elements of a model in computer language. However, just expressing a social mechanism in computer code is not enough.

Good practice #1: Formulate precise descriptions of the social phenomenon to be generated in terms of the macro-level outcome, its situational context and of the postulated generative mechanism before writing computer code.

Good practice #1 helps model builders keep their focus on the research problem and the theoretical assumptions that constitute a postulated mechanism rather than on technical details of their computational implementation.

> *Example of a research goal in ABCM:*
>
> Show that a relocation cascade can generate higher levels of segregation than needed to satisfy individuals' ethnic preferences. Further, identify the conditions under which this outcome obtains as robust regularity.

The macro-level outcome of interest is here specified as segregation, the proposed explanatory mechanism is that of a relocation cascade and a key element of the situational context is that ethnic preferences are mild in the sense that individuals would be content with less segregation than they get. The *NetLogo* model *SegregationExtended* gives specific technical elaborations for each of these core substantive concepts.

3.2 A Slightly More Realistic Toy-Model

SegregationExtended encompasses Schelling's classic toy-model as a special case.[5] We start here from a version with slightly more realistic assumptions, assuming in particular larger neighborhoods and a population consisting of four groups roughly resembling the ethnic composition of a contemporary large US city. Agents in the program are called *households*. The **model of agent behavior** is extremely simple. Households' only "action" is their choice of a residential location, represented by which cell of a rectangular checkerboard-like "world" (hereafter: cellular world) the household chooses to locate to at a given point of time. Further, households have the state variable *ethnicity*, assigned as a fixed characteristic upon initialization of the simulation.

[5] This chapter cannot give an introduction to *NetLogo* itself. Further explanation of the model can be found under the tabs "Code" and "Info", respectively, on the model's interface and in many explanatory comments in the code added by the authors (starting with ";;"). The code of the models *SegregationExtended* and *SegregationDiscreteChoice* has not been optimized for efficiency, but instead for transparency for users new to NetLogo. A tutorial on the NetLogo website (https://ccl.northwestern.edu/netlogo/docs/) provides an accessible introduction to NetLogo itself. Further comprehensive introductions in the literature are Wilensky and Rand (2015) and Railsback and Grimm (2019).

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

By default, *SegregationExtended* assumes a **situational context** with four different groups (*ethnicity* ∈ {1,2,3,4}). Following Clark and Fossett (2008), the agent population resembles in this setting roughly the composition of a typical large US city with a white majority, and African Americans, Asians, and Hispanics with proportions of 60%, 20%, 10%, and 10%, respectively. In the model, these groups are called Red, Green, Blue, and Yellow, respectively. Switching *fourGroups* "Off" creates two equally large groups Red and Green (∈ {1,2}), as often used in the literature (Hedström and Bearman, 2009). In both settings, households are in the initial condition randomly distributed over the cellular world. While this may be empirically implausible given the high levels of ethnic segregation observed in many multi-ethnic societies, this initial condition is of high theoretical interest for assessing whether and under which conditions relocation cascades give rise to high levels of segregation, especially if households do not prefer nor live in segregated neighborhoods initially.

Figure 24.1 shows examples of states of the cellular world with four groups as they are shown on the *NetLogo* interface of *SegregationExtended*. The probability that a cell is occupied by a household rather than being empty is a further important element of the situational context in this model, defined by the program slider *density*.

Households' ethnic preferences are represented by the threshold *T* for the minimal proportion of co-ethnics a household wants to have (slider *%-similar-wanted*). In *SegregationExtended*, *T* is measured in percentage points and equal for all households, represented by a so-called "global" variable known to all households in the program. The model makes exactly one simple assumption about households' cognition (see procedure *update-households*). If the proportion of co-ethnic households in the neighborhood falls below *T*, the household is unhappy. The model of action is equally simple (see procedure *move-unhappy-households*). If a household is unhappy, it searches randomly a vacant spot, starting from its current location, and stays in it as soon it has found one.

Figure 24.2C shows how households are **locally interdependent** with their neighbors. Typically, Schelling's (1971) model is presented with the assumption of very small neighborhoods with up to eight local neighbors (see also Hedström and Bearman, 2009), but this may seem empirically implausible. Benenson et al., (2009), for example, propose that neighborhoods encompass the houses around a residential location that are visually noticeable. Depending on the exact spatial structure, this could be considerably more than eight households. Thus, the model assumes by default a larger neighborhood size with up to 28 neighbors (see Figure 24.2C). Yet, model assumptions about neighborhood size may vary, depending on theoretical and empirical considerations. The slider *radius* in *SegregationExtended* allows changing neighborhoods as demonstrated in Figure 24.2C, where *radius* = 3 is the default. For households located on the border of the cellular world, their neighborhood does not stretch beyond the border.[6]

The button *Go* on the NetLogo interface starts the procedure *go*, which controls the simulation execution. In every round of the simulation (*tick* in NetLogo), the procedure *move-unhappy-households* calls on all households in a random sequence and activates the procedure *find-new-spot* for unhappy households. Further, the model contains a very simple **stopping rule**. When all agents are happy the simulation ends, because an equilibrium state has been reached in which no further relocation can occur.

[6] Often it is assumed that cellular worlds are "wrapped around" on a torus (Hegselmann and Flache, 1998) to avoid border-effects. Our world here has borders, just as real cities have.

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

Finally, the model reports **outcome measures** and graphical output (e.g. Figure 24.2A) to show how **macro-level characteristics** change when the micro-level processes of household relocation dynamics are at work. *Percent Similar*, the average proportion of co-ethnics among a household's neighbors, is computed for the entire population and separately for each of the four groups. Comparison of its value over time with the initial level shows segregation, indicating whether households sort into clusters that are more ethnically homogenous than in the initial randomly mixed population. Also, a comparison of its value with the threshold *T* tells whether households end up more segregated than needed to satisfy their ethnic preference. Further outcome measures will be introduced next.

3.3 Model Analysis: Dissecting a Toy-Model

Hedström and Bearman (2009, p. 13) demonstrated how the simplest version of Schelling's model can generate a complex micro-macro interaction. Using two equally sized groups and a small neighborhood, they showed how an increase of the ethnic preference *T* by only one percentage point (from 25% to 26%) remarkably changed the macro-level outcome from near random mixing to clear segregation. *SegregationExtended* slightly differs in some model details from their version, for example in the exact shape of neighborhoods or the exact number of households and density.[7] Yet, the reader can easily check that *SegregationExtended* produces almost the same result by running simulations for a situational context similar to theirs (*radius* = 1, *fourGroups* = *Off*, 51x51 world size, 90% density). In other words, the mechanism of relocation cascades meets our research goal independently from some details of how it is programmed. This demonstrates an important good practice in ABCM.

Good practice #2: Always test whether substantively important model results are robust against small changes of theoretically irrelevant model details.

But will we find a similarly sharp transition in macro-level outcomes also for our more realistic default scenario with four groups and larger neighborhoods? Figure 24.1 shows the outcomes of a conservative test. We compared model results for $T = 25$ and $T = 35$. The figure shows two representative outcomes that emerged after 1000 simulation rounds with the default settings of *SegregationExtended*.

As it turns out, the more realistic model cannot generate a similarly sharp qualitative transition in macro outcomes. Figure 24.1 shows that an increase of ethnic preference *T* by 10% yields an increase of the level of segregation by about 4, a change that is very modest compared to the strong effect of a 1% increase of *T* when we assume two groups, radius = 1 and otherwise the same conditions ($T = 25$ and $T = 26$ yielded 58% and 78.7%, respectively). Further simulations with our more realistic scenario show that for $T = 25$ vs. $T = 26$, the difference in segregation is now almost negligible, with *Percent Similar* of 52.8 vs. 53.1 in two representative runs.

[7] The situational conditions in their simulation differ from the standard settings in *SegregationExtended* as follows. They use a cellular world without borders, they assume exactly 1 250 households per group and they use a so-called Moore neighborhood with 8 neighbors (NetLogo dictionary: *neighbors*), while our *radius* = 1 imposes four neighbors (Von Neumann neighborhood). Nevertheless, outcomes are very similar. When we increased *T* from 25 to 26, *Percent Similar* soared from 55.6 to 77.3 in two representative runs with *SegregationExtended*.

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

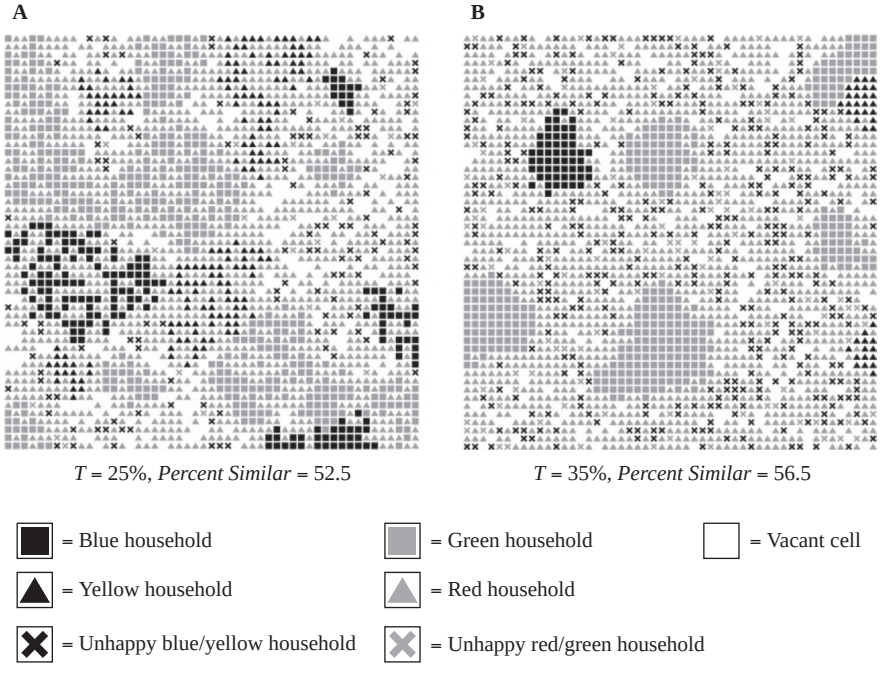


Figure 24.1 Final states of two representative runs for different values of ethnic preference threshold (T). The proportion of Red, Green, Blue, and Yellow households is 60%, 20%, 10% and 10%, respectively. Groups are based on Clark and Fossett (2008): African American = Green; Asian = Blue; Hispanic = Yellow; White = Red. World size = 51x51, 90% density (≈2345 households and ≈255 empty cells)

This discrepancy between the original toy-model and our slightly more realistic version serves to illustrate several important good practices in ABCM.

Good practice #3. Show the statistical distribution of model outcomes.

Good practice #4. Always formulate a candidate explanation and tentative conjecture about the expected model outcomes before running a simulation experiment.

Good practice #5. Design and conduct simulation experiments for testing the candidate explanation.

Good practice #3 is straightforward and common in ABCM research. Models typically generate stochastic processes and single runs may say little about the distribution of results. To assure robustness, we used NetLogo's *BehaviorSpace* feature[8] to run 10 independent repetitions per condition. Results supported robustness, yielding averages and standard deviations for *Percent Similar* in the final state of 53.27 (0.63), 53.04 (1.21) and 56.32 (2.89) for $T = 25$,

[8] The implementations of all simulation experiments shown in this chapter are embedded in the models *SegregationExtended* and *SegregationDiscreteChoice* and can be found under Tools > BehaviorSpace on the NetLogo interface.

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

26 and 35, respectively. In all examples shown in this chapter, there was very little variation between the outcome measures of different independent repetitions, which made further statistical tests or visualization of distributions unnecessary. This is not always the case and sometimes many more repetitions may be required.

Good practices #4 and #5 are less straightforward. However impressively an ABCM demonstrates a complex micro-macro interaction, modelers must be able to explain why. "The computer said so" is not good enough as an answer. Beautiful results from simulation studies may turn out to be hard to communicate to other researchers or practitioners and, what is more, there is no a priori guarantee that simulation outcomes are systematically generated by the theoretical mechanism the modeler wanted to implement rather than by inessential technical details or—even worse—programming errors. Every experienced ABCM practitioner knows stories of how beautifully counterintuitive insights from simulation models fell apart after bugs in the code or conceptual artefacts in the model were discovered (see, for example, a discussion in the literature about an extension of Schelling's model: Bruch and Mare, 2006; van de Rijt et al., 2009).

It is difficult to pin down what exactly explaining model outcomes means. Ideally, mathematical proofs can be given that a simulation algorithm or a simpler formalization of essentially the same mechanism must generate the observed outcomes. Sometimes such proofs can be found or are available in the literature, e.g. for the simplest versions of Schelling's model (Young, 1998). If they are, they form an important source of guidance for model understanding. However, often such results are simply not available (yet) or only apply for overly strong simplifications of a model. Even then, researchers should be able to produce clear causal arguments for why a model generates a certain outcome under certain conditions. Similar to theory-driven empirical research, the ultimate test is to derive expectations about model behavior in yet unexplored conditions and test these expectations against the computational evidence the model generates. Unlike empirical research, ABCM allows modelers to experiment with and fully observe every detail of the complex system the researcher created. This allows designing and running experiments which advance understanding of the model by putting tentative explanations to a critical test.

Let us apply good practices #4 and #5 to our example. Why at all does segregation increase if we raise *T*, and why, more in particular, is the micro-macro interaction of our more realistic model so much less dramatic than for the simpler version with two equally sized groups and *radius* = 1? Answering the first question provides a candidate explanation that helps to answer the second. The larger the proportion of initially unhappy households, the more households will move and the bigger the relocation cascades eventually will become. Further, the bigger the relocation cascades are, the more segregation arises in the final state, because every relocation tends to increase neighborhoods' ethnic homogeneity. This yields a testable expectation. The larger the proportion of initially unhappy households, the higher segregation will eventually become. As the next experiment will show, this reveals why the micro-macro interaction is qualitatively different to the original toy-model. The crucial reason is that neighborhoods are much larger with *radius* = 3.

Figures 24.2A and B show results of a simulation experiment (Experiment 1 and 2 in *BehaviorSpace*, respectively) designed to systematically test the candidate explanation. Ethnic preference *T* was varied starting from *T* = 0% to *T* = 70% in steps of 1%, and the experiments were conducted for two different neighborhood sizes (*radius* 1 vs. 3). Simulations were stopped after maximally 1 000 rounds if they did not reach equilibrium before. The chart reports two outcome measures, the proportion of households who were unhappy in the initial

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

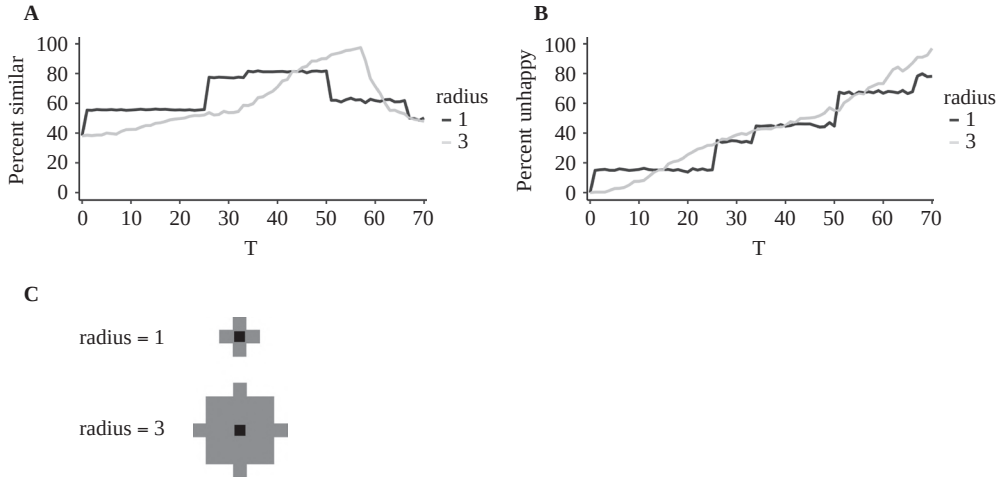


Figure 24.2 Effects of ethnic preferences (T) on percent similar after 1000 simulation rounds (A) and percent unhappy initially (B), averaged across 10 repetitions per level of T. Neighborhoods with radius 1 and 3 are visualized in panel C. Panels A and B are generated with Experiment 1 and 2, respectively. World-size 51x51, 90% density (about 2345 households)

state (Experiment 2) and *Percent Similar* in the final state (Experiment 1). The addition of the initial proportion of unhappy households illustrates that in combination with systematic experimentation, properly designed outcome measures collecting model-generated "data" about the internal processes of a model are an important source of information for testing tentative explanations of model outcomes.

Results clearly support our expectations as long as *T* is not too high. Especially the results for *radius* = 1 show that the levels of *T* at which the initial proportion of unhappy agents rises sharply, coincide exactly with similarly sharp increases of *Percent Similar*. For *radius* = 3 we see instead a much smoother increase, alongside a similarly smooth increase of the initial proportion of unhappy households.

Experiments 1 and 2 partially support our candidate explanation. But why are outcomes so different for *radius* = 1 and *radius* = 3? The difference in neighborhood structures (Figure 24.2C) shows why. Households with four neighbors (radius 1) want at least one co-ethnic if $0 < T <= 25\%$, they require at least two co-ethnics if $25\% < T <= 50\%$, three co-ethnics are needed when $50\% < T <= 75\%$ and only four co-ethnics can satisfy a household when $T > 75\%$. This is why under *radius* = 1 the proportion of initially unhappy households goes upward in sudden jumps when ethnic preferences *T* change, inducing corresponding sudden jumps in segregation.[9] However illustrative, the strong micro-macro interaction demonstrated for *radius* = 1 turns out to be an artefact caused by extremely small neighborhoods. With the larger neighborhood of *radius* = 3, a household's initial neighbors resemble a representative sample of the overall population. Here, increasing

[9] Figure 24.2A also shows smaller jumps for *radius* = 1 at *T* = 33 and *T* = 66. The reason is that households at the border have three rather than four neighbors.

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

T changes initial happiness rates much more smoothly, proportionally to how likely an average household has *T*% or more co-ethnic neighbors in a random sample (N = 28) of the overall population.

Putting good practices #4 and #5 into action, Experiments 1 and 2 tested our candidate explanation. However, this test yielded only partial confirmation. What was not anticipated was that segregation in Figure 24.2A suddenly dropped beyond a critical level of *T*, although initial unhappiness further increased (Figure 24.2B). For *radius* = 1, the drop occurs at about T = 50%, for *radius* =3 the critical point is at about T = 60%. This brings us to the next good practice.

Good practice #6. Inspect single runs of the conditions where model behavior cannot yet be explained.

Observing the dynamics of single runs at various levels of T > 50% quickly shows what happened. It is nearly impossible that all households find vacant locations with this many co-ethnic neighbors just by a random walk. Thus, even after 1 000 simulation rounds many households remain unhappy and keep wandering around at random, strongly reducing segregation. Majority households can still form clusters to some extent, but minority households fail to do so. This explains why *Percent Similar* does not fully drop to the level of random mixing. Our candidate explanation needs refinement. If *T* exceeds a critical level, the proportion of unhappy households becomes too high so that the formation of stable ethnic clusters becomes impossible and segregation drops.

The example illustrates two lessons of more general importance. First, it shows how important it is to make sure one has understood unexpected model results before further complications are added to a model.

Good practice #7. Always develop an explanation of unexpected simulation results and test it, before developing the model further.

The example also demonstrates how important it is to think carefully about the stopping rule applied in a simulation experiment and to explore model behavior with single runs a priori. Had we not fixed the maximum number of rounds at 1 000, Experiment 1 would never have finished within a reasonable amount of time. That is, the other stopping condition—that all agents are happy—was practically impossible to ever be met for high levels of *T*. Inspecting single runs before actually experimenting made it possible to find this out in advance.

Good practice #8. Inspect single runs by sampling the whole parameter space of an experiment before designing the experiment and stopping rule.

On the whole, the somewhat more realistic model we have presented here meets the research goal formulated. We identified conditions for which it generates levels of segregation which clearly exceed those of random mixing and those needed to satisfy agents' preferences *T*. At the same time, our model differs in even more ways from Schelling's original toy-model than could be discussed here. This includes, in particular, the counterintuitive result that some groups actually cluster less when their ethnic preferences become more demanding. The explanation for this also makes clear why in Figure 24.2A we observe some decline of *Percent Similar* between about T = 25 and T = 30, another result not anticipated by our a priori

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

candidate explanation. The reader can find elsewhere how this intriguing complexity can be discovered and unraveled with further simulation experiments.[10]

3.4 Toward a More Realistic Model of Relocation Decision-Making

3.4.1 Model extension

In this section, a final model extension is presented in which we changed three aspects of the model of **agent behavior**, increasing the empirical plausibility of assumptions about relocation decision-making. In a real ABCM project these aspects should be introduced and analyzed step by step. For lack of space, we focus here on assessing whether the resulting model meets our research goal.[11]

First, ethnic preferences are made heterogeneous in that households differ within and between groups in how tolerant they are toward outgroups in their neighborhood (Xie and Zhou, 2012). Second, we drop the assumption of thresholds, assuming instead that neighborhood attractiveness differs depending on how closely households' ideal neighborhood composition is matched (Zhang, 2004). Finally, the more attractive a vacant location is for a household, the more likely the household will relocate to it (Bruch and Mare, 2006; van de Rijt et al., 2009). This last extension, in particular, connects ABCM with empirical research (Bruch and Atwell, 2015), using discrete choice models (Bruch and Mare, 2009) for which the parameters can be estimated from empirical data (van Gent et al., 2019), including experiments (Holm et al., 2016).

Ethnic preferences in *SegregationDiscreteChoice* are loosely based on empirical data. Table 24.1 summarizes neighborhood preferences as elicited in a US survey study with approximately 1,000 respondents per major ethnic group (Clark and Fossett, 2008). The data show the proportion of respondents per group who found a certain number of co-ethnics ideal in a hypothetical neighborhood containing eight households.

SegregationDiscreteChoice models these preferences with single-peaked neighborhood attractiveness functions. More precisely, for a household with ideal share T, the attractiveness A of a neighborhood with a real share s of co-ethnic households is computed as in Equation (24.1):

$$A(s,T) = \begin{cases} \dfrac{s}{T} & \text{if } s \le T \\ 1-(s-T) & \text{if } s > T \end{cases} \tag{24.1}$$

Figure 24.3 illustrates the substantive assumptions implemented by Equation (24.1). Households find a neighborhood with their ideal composition maximally attractive (A = 1). The more a neighborhood deviates from this ideal, the less attractive it is. Further, neighborhoods without

[10] For more information, we refer to our research cluster website where we discuss an extension considering the impact of T on clustering per ethnic group as well as varying the level of density between 70% and 90%: https://sites.google.com/view/normsandnetworks/addendum-chapter-agent-based-computational-models-in-analytical-sociology.

[11] The model presented in this section was implemented in the separate *NetLogo* program *SegregationDiscreteChoice*.

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

Table 24.1 Proportions of respondents of four different ethnic groups in the US (African American = Green; Asian = Blue; Hispanic = Yellow; White = Red), choosing an ideal neighborhood with a given number of co-ethnic households (maximum 8). The final column (Modal [T]) indicates the proportion of co-ethnics corresponding with the modal scenario in a category; the left column visualizes a corresponding scenario. The center cell represents the household, and grey and white cells represent same-group and other-group neighbors, respectively. For more details, please see Clark and Fossett (2008, p.4111)

Tolerance: Number of co-ethnics preferred (out of 8)			Green (20%)	Blue (10%)	Yellow (10%)	Red (60%)	Modal (T)
7/8		*Low tolerance (8–6)*	11.5	22.6	19.6	22.3	87.5
4/8		*Medium tolerance (5–3)*	49	51.9	45.9	48.7	50
1/8		*High tolerance (2–0)*	39.5	25.5	34.5	29	12.5
		Total	100	100	100	100	

co-ethnic neighbors are maximally unattractive ($A = 0$), and the higher a household's ideal proportion, the more attractive it deems purely co-ethnic neighborhoods ($A(1,T) = T$). *SegregationDiscreteChoice* models households' decision-making following the standard form of discrete choice models as given by Equation (24.2):

$$P_{new} = \frac{e^{\beta A(s_{new},T)}}{e^{\beta A(s_{new},T)} + e^{\beta A(s_{old},T)}} \qquad (24.2)^{12}$$

In Equation (24.2), the probability that the new location will be chosen is P_{new}, and the symbols s_{new} and s_{old} indicate the proportion co-ethnics in the potential new and the old neighborhood, respectively. The parameter β scales how closely the choice probability is linked with the relative attractiveness of the two neighborhoods (see procedure *betaDiscreteChoice*). The higher β, the less likely a household chooses the less attractive neighborhood.

The probabilistic discrete choice model can be considered empirically more plausible than the original threshold function. In particular, whether a household leaves a neighborhood for a new one is no longer fully determined by these neighborhoods' compositions, but is also due to random chance. Implicitly, this reflects that in reality households' decisions are also shaped by factors other than neighborhood composition. This model modification illustrates a further important good practice in ABCM.

Good practice #9. Test whether model results of substantive importance are robust if small amounts of random deviations are added to deterministically formulated model assumptions.

[12] In *SegregationDiscreteChoice*, this function is implemented in a mathematically equivalent but computationally more efficient form as $P_{new} = \frac{1}{1+e^{\beta(A(s_{old},T)-A(s_{new},T))}}$.

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

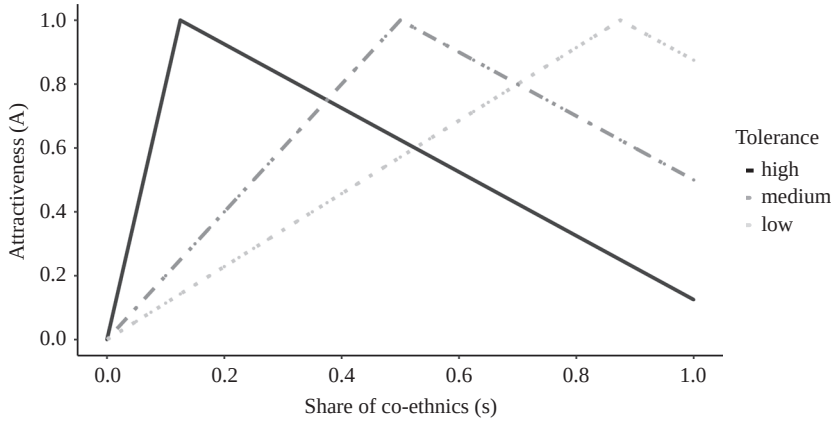


Figure 24.3 Single-peaked ethnic preference functions for households in three different tolerance categories (based on Table 24.1). Attractiveness reflects how attractive the share of co-ethnics (s) is. Note that a low (high) tolerance means that many (few) co-ethnics are wanted

In *SegregationExtended*, an unhappy household always abandoned a neighborhood and a happy household never did. This deterministic assumption is relaxed in *SegregationDiscreteChoice*. In a real ABCM project, nothing else should be changed when randomness is added to a model for evaluating the robustness of central results. For lack of space, we cannot demonstrate a corresponding test here. However, *SegregationExtended* contains a slider *noise* that allows one to explore whether results change if a small probability of random perturbation of households' happiness is added. Readers are invited to conduct this test and search for explanations if they find that model outcomes change. Generally, testing effects of adding small amounts of random noise is very important in ABCM research as this may not only profoundly change model results, but sometimes also improve the empirical accuracy of model predictions (Macy and Tsvetkova, 2015; Mäs and Helbing, 2020).

The discrete choice model also requires a new method for deciding which households get a chance to relocate. *SegregationDiscreteChoice* adopts the simple solution to select every household with a certain probability per round, given by the parameter *probMoveOption*. Further, the procedure *update-households* selects at random who can move in the next round, and the procedures *move-unhappy-households* and *find-new-spot* now contain code for memorizing the origin location and its attractiveness before the random walk. This allows moving a household back to its original location if it decides to not relocate. Finally, new procedures *attractiveness* and *BinaryChoice* compute *A(s,T)* for a given neighborhood and conduct the random choice based on Equation (24.2).

The procedure *setup* in *SegregationDiscreteChoice* randomly distributes ethnic preferences *T* across households based on the relative frequencies of household categories in Table 24.1. Other than in *SegregationExtended*, ethnic preference *T* (*%-similar-wanted*) is no longer a global variable but a fixed internal state of an individual household agent (*my-%-similar-wanted*).

3.4.2 Exploring implications

Arguably, *SegregationDiscreteChoice* is more realistic than *SegregationExtended*. But why add all this realism? Manzo (2014) points to an important reason. "Injecting realism" tests whether a model's ability to provide a generative explanation holds up under further empirically motivated constraints on the mechanism.

For our example, this means we need to test whether *SegregationDiscreteChoice* still meets the research goal and generates segregation higher than would be needed for satisfying households' ethnic preferences. With single-peaked preferences, this requires new outcome measures. *SegregationDiscreteChoice* computes a set of outcome measures called *oversegregation*, reporting for every household category the average of the difference between the desired share of co-ethnic neighbors (*T*) and the actual share obtained.

Adding new features to a model also offers researchers the opportunity to formulate and test expectations about effects. In our case, we could ask how a category's ethnic preference is related to how much it clusters with co-ethnics in the emergent residential distribution. We invite readers to formulate a hypothetical answer based on current model understanding. To test answers, a new measure for spatial *clustering* divides for every household category its average *Percent Similar* by the ethnic group's population share. Values above 1 indicate that this type of household forms spatial clusters which are ethnically more homogenous than would be expected under random mixing.

Figure 24.4 shows results of a single run (A), and Experiment 3 (B) provides tentative answers. Experiment 3 combines the *world size*, *density*, *radius*, and *ethnic composition* settings of Experiments 1 and 2 with the heterogeneous ethnic preferences based on Table 24.1. Further, *probMoveOption* = 5% approximates moving rates in real cities (Benenson et al., 2009), and β = 12 sets relative neighborhood attractiveness to be the dominant factor in household's choices, but leaves room for some random variation.

Figure 24.4A shows a single run charting the evolution of clustering for the three different subcategories of the majority group Red and minority Blue over 2 000 rounds. First, we can see that *SegregationDiscreteChoice* still generates the emergence of segregated neighborhoods from initial integration. Except for tolerant majority households, all groups become more clustered than under random mixing. At the same time, the increase of clustering is only steep for medium and high threshold minority groups, while for all other groups clustering increases

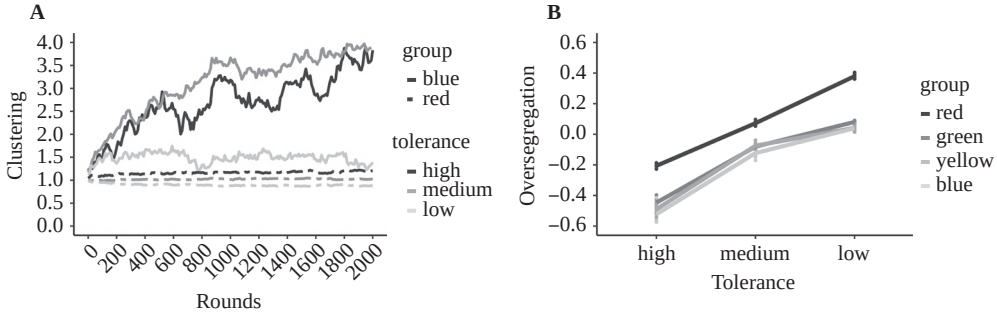


Figure 24.4 Panel A: Change of clustering for two selected households (Blue and Red) categories in a single run over 2 000 rounds. Panel B: Variance and average oversegregation (10 repetitions) by tolerance category after 1 000 rounds

only slightly. Further, we see that tolerance affects residential outcomes. In both ethnic groups, the most tolerant households end up in less segregated neighborhoods than their less tolerant counterparts. However, the precise link between (in)tolerance and clustering is more complex.

Figure 24.4B shows results for the same condition, explicating how ethnic preferences relate to households' average *oversegregation* after 1 000 rounds (Experiment 3). There are only two types of households whose desired neighborhood composition clearly deviates from what they get: tolerant majority households and intolerant minority households. Tolerant majority households have many more co-ethnic neighbors than they would like; intolerant minority households have far too few.

Further refinement of our tentative theory of relocation cascades is required. Based on approximately realistic assumptions about ethnic preferences, the mechanism implemented in *SegregationDiscreteChoice* results only for some types of households in higher segregation than they desire, and for some it even generates less. Again, in a real ABCM research project, further carefully designed experiments would be needed to find out why.

We conclude this example by pointing to an intriguing experiment the reader can conduct with *SegregationDiscreteChoice* in search for an answer. Inspection of single runs with the standard scenario suggests a reason why intolerant minority households fail to become as clustered as they would like. As soon as minority clusters form, vacant spots in them are "invaded" by tolerant majority households seeking to reside in multi-ethnic neighborhoods. Tolerant Red households comprise about 18% of the whole population, more than each of the two minorities, Blue and Yellow. As a consequence, for the taste of intolerant Blue and Yellow households, too many Red households occupy vacant spots in "their" emergent ethnic clusters, driving them away. This explains the relatively low degree of ethnic clustering of intolerant minorities, compared to their medium-tolerant co-ethnics. The explanation can be tested with *SegregationDiscreteChoice*. The model interface comprises a switch *TolerantMajority*, which is set "On" by default. With *TolerantMajority* switched "Off", all formerly tolerant households in the majority group are turned into medium-tolerant when the model is initialized. Nothing else is changed. A ceteris-paribus replication of Experiment 3 with *TolerantMajority = Off* confirms the explanation. While the over-segregation index was about –0.5 for these groups in Experiment 3, it rises to about –0.3 and –0.2 for intolerant Blue and Yellow households, respectively. Visual inspection reveals a pattern not visible before. Now, much denser minority clusters emerge, which tend to develop a core-periphery structure in which intolerant households form the center surrounded by more tolerant co-ethnics.

This concluding example underlines two important insights. First, it illustrates (again) that making a model less realistic can be a good practice for better understanding consequences of adding empirically more plausible assumptions. While the empirical data suggest that the household population should contain Red households with highly tolerant preferences, dynamics of the model can be better understood if we remove this category for a moment.

Good practice #10. Move back to analyze the behavior of a model for less realistic conditions if needed to test or develop explanations of model behavior for a more realistic condition.

More central to the program of AS, the example also illustrates how misleading it can be to infer explanations for complex social phenomena from observed individual-level behavior alone. If this were the real world, a researcher could think that strongly segregated core-

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

periphery neighborhoods of ethnic minorities are caused by these minorities' xenophobic ethnic preferences. As we found, this explanation would be incomplete at best. Minorities with the same preference structure can end up in much less segregated neighborhoods if more majority households in the population are tolerant. Complex interdependencies like this one occur in many social mechanisms. ABCM is a powerful tool to unravel them.

3.5 Even More Realism in Models of Segregation

The simulation experiments shown here illustrate 10 good practices for ABCM. Moreover, they illustrate how ABCM is useful for AS by adding in a step-like fashion realism to a model. Yet, *SegregationDiscreteChoice* could still be deemed a toy-model neglecting numerous factors known to affect empirical residential segregation. It would be misguided to see this as a limitation of ABCM as such. ABCMs of residential segregation were developed that start from models similar to those shown here, incorporating housing prices, income differences, neighborhood reputation (Zhang, 2004; Fossett, 2006), or the real spatial arrangement of a city based on GIS data (Benenson et al., 2009) as further elements.

4. CONCLUSION AND DISCUSSION

ABCM plays a central role in analytical sociology. Our chapter focuses on what we see as its most important strength, the ability to dissect with systematic experimentation and model analysis a potentially complex theoretically proposed social mechanism. This is especially important because full comprehension of an empirically realistic model can be a formidable challenge. Heterogeneity of an agent population, inherent randomness, complex network structures, or local and non-linear interdependencies between multiple agents are but a few of the features many real social systems exhibit. The complexity of social reality is why social scientists increasingly resort to formal modeling and ABCM in the first place (Mäs, Chapter 4, this volume). The gain is that an ABCM is simpler and much more amenable for scientific analysis than social reality itself. Yet, a reasonably empirically plausible model can still be a complex system in itself where the relation between model assumptions and outcomes is far from trivial. A full understanding of why, how, and under which conditions a model is capable of generating outcomes resembling a real social phenomenon is thus a precondition that needs to be met before the ABCM can be considered to provide a valid explanation of the empirical phenomenon it targets.

The good practices we proposed help assure that the strengths of ABCM can be put to its best use for Analytical Sociologists. Our good practices are part of a broader set of established methodologies and principles in ABCM, which are explained and discussed in comprehensive introductions such as (Squazzoni, 2012; Edmonds and Meyer, 2017; Railsback and Grimm, 2019). In addition to model understanding, they also stress the importance of correct implementation of model assumptions (verification), proper use of empirical data for choosing initial conditions and parameter values (calibration), and appropriate testing of model implications against empirical evidence (validation). As the field is moving quickly, interested readers are also advised to consult journals on the intersection of ABCM and sociology, such as *The Journal of Artificial Societies and Social Simulation (JASSS)*, *Computational and Mathematical Organization Theory (CMOT)* or the *Journal of Mathematical Sociology*

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

(JMS). Readers should further be aware that it has become a generally accepted practice that computer code of published models is made accessible for replication (e.g. CoMSES Net/OpenABM, 2020), following standard protocols such as ODD (Grimm et al., 2020). Debates around the lack of transparency of highly influential models such as those used for predicting mortality and infection rates in the recent COVID-19 crisis have highlighted the special importance of this particular practice (Squazzoni et al., 2020). Finally, one should realize that beyond *NetLogo* there is a multitude of software packages and programming languages available for ABCM. Which of these is the best for a given simulation project depends on many factors, including programming experience of the researchers, available software and hardware, or need for large-scale time-intensive computational experiments. Reviews of the field help to navigate this complex landscape (Abar et al., 2017).

We conclude by reflecting on three debates which are of particular relevance for AS concerning the relation between toy-models and empirically more realistic models, explanation vs. theoretical exploration as research goals, and causal inferences drawn from ABCM.

First, AS rejects the view that empirically unrealistic models should be preferred for considerations of higher elegance or tractability. Yet, this does not mean that model builders have nothing to learn from knowing and understanding "toy-models" available in the literature (Reutlinger et al., 2017). Often, more general mathematical analyses are available or can be constructed for these models, which help to assess and understand whether and under which conditions more detail and realism really change fundamental properties of a model (see, Keijzer et al., 2018, for a recent example). Furthermore, unrealistically simple models can shed light on why seemingly plausible explanations do not work as expected, or highlight new potential explanations in a maximally lucid way, as Schelling's original model did. Finally, in many cases realistic mechanistic explanations combine elements of several simpler, well-understood mechanisms. For example, while relocation cascades can be one part of the explanation on why segregation arises in a real city, another interconnected mechanism might be increasing opinion polarization between segregated groups, driven by ethnic homophily and social influence processes (Axelrod, 1997; Mäs et al., 2013; Feliciani et al., 2017; Flache, 2018, 2019). Add to this the very real possibility that further processes interfere, such as housing market dynamics, or emergent neighborhood reputations driven by self-reinforcing social influence (van de Rijt, 2019), and it should be clear that researchers are well advised to first get a solid grasp of the dynamics of each of these mechanisms in isolation before attempting to increase realism by integrating them in one model.

Second, a related question is whether ABCM should exclusively be used for providing explanations of empirically observed social phenomena. We believe that this would leave an important potential of ABCM unused. New and fruitful theoretical ideas can emerge from simulation projects asking the "what if" question, focusing on possible but potentially as-yet unobserved consequences of fundamental social processes which are empirically plausible, but are connected in a theoretically new way. ABCM can help to discover and unravel such consequences, which then can give rise to new hypotheses that are testable in experiments or field studies. One example is the insight that a cohesive social network may undermine rather than foster cooperation in the provision of collective goods. This was first highlighted by a purely theoretical ABCM work (Flache and Macy, 1996), which then informed subsequent experimental tests (Flache, 1996; Flache et al., 2017) and resonated later findings from independently conducted field research (Langfred, 1998). Similarly, in studies of opinion polarization, new insights about the role of demographic faultlines were first explored theoretically in computational experiments and then tested in lab experiments (Mäs and Flache, 2013), as was the

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

insight that complex contagion can play a key role in the diffusion of behaviors in clustered social networks (Centola, 2010), which has inspired further ABCM research showing how incorporation of empirical data on actual social influence processes and network structures can yield new conditions and explain empirical variation in adoption curves between different communities (Manzo et al., 2018). Related to this point, there always is the possibility of computational serendipity. Models designed to explain a specific phenomenon can turn out to generate unexpected implications, which—after a closer look—appear empirically plausible and could inspire new research.

Finally, a key issue in AS is whether an ABCM can actually show that the mechanism it implements really causes the empirical phenomenon it targets. In short, the answer given by analytical sociologists (Hedström and Ylikoski, 2010; Manzo, 2021) is that the more the elements that are constituting a model, e.g. its agents and behavioral and structural assumptions mirroring empirical evidence about them, the more plausible is the claim that the mechanism embodied in the model also causes the phenomenon if the model succeeds in generating macro-outcomes resembling this phenomenon.

Handling the tool of ABCM properly requires adherence to a set of good practices that assure that researchers fully comprehend the complexity of their own theoretical creations. Under this precondition, ABCM is a unique method for providing precise, causal and empirically plausible mechanistic explanations of puzzling phenomena in the social world.

REFERENCES

Abar, S., G.K. Theodoropoulos, P. Lemarinier and G.M.P. O'Hare (2017), "Agent based modelling and simulation tools: A review of the state-of-art software", *Computer Science Review*, **24**: 13–33. doi: 10.1016/j.cosrev.2017.03.001.

Axelrod, R. (1997), "The dissemination of culture: A model with local convergence and global polarization", *Journal of Conflict Resolution*, **41** (2): 203–26. doi: 10.1177/0022002797041002001.

Baldassarri, D. and P. Bearman (2007), "Dynamics of political polarization", *American Sociological Review*, **72** (5): 784–811. doi: 10.1177/000312240707200507.

Bendor, J. and P. Swistak (2001), "The evolution of norms", *American Journal of Sociology*, **10** (6): 1493–545. doi: 10.1086/321298.

Benenson, I., E. Hatna and E. Or (2009), "From Schelling to spatially explicit modeling of urban ethnic and economic residential dynamics", *Sociological Methods and Research*, **37** (4): 463–97. doi: 10.1177/0049124109334792.

Bianchi, F., A. Flache and F. Squazzoni (2020), "Solidarity in collaboration networks when everyone competes for the strongest partner: A stochastic actor-based simulation model", *Journal of Mathematical Sociology*, **44** (4): 249–66. doi: 10.1080/0022250X.2019.1704284.

Bianchi, F. and F. Squazzoni (2015), "Agent-based models in sociology", *Wiley Interdisciplinary Reviews: Computational Statistics*, **7** (4): 284–306. doi: 10.1002/wics.1356.

Breen, R. (2009), "Game theory", in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press, pp. 619–38.

Bruch, E.E. and J. Atwell (2015), "Agent-based models in empirical social research", *Sociological Methods and Research*, **44** (2): 186–221. doi: 10.1177/0049124113506405.

Bruch, E.E. and R.D. Mare (2006), "Neighborhood choice and neighborhood change", *American Journal of Sociology*, **112** (3): 667–709. doi: 10.1086/507856.

Bruch, E.E. and R.D. Mare (2009), "Segregation dynamics", in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press, pp. 269–93. doi: 10.1093/oxfordhb/9780199215362.013.12.

Castellano, C., S. Fortunato and V. Loreto (2009), "Statistical physics of social dynamics", *Reviews of Modern Physics*, **81** (2): 591–646. doi: 10.1103/RevModPhys.81.591.

Centola, D. (2010), "The spread of behavior in an online social network experiment", *Science*, **329** (5996): 1194–7. doi: 10.1126/science.1185231.

Centola, D., J.C. González-Avella, V.M. Eguíluz and M. San Miguel. (2007), "Homophily, cultural drift, and the co-evolution of cultural groups", *Journal of Conflict Resolution*, **51** (6): 905–29. doi: 10.1177/0022002707307632.

Clark, W.A.V. and M. Fossett (2008), "Understanding the social context of the Schelling segregation model", *Proceedings of the National Academy of Sciences of the United States of America*, **105** (11): 4109–14. doi: 10.1073/pnas.0708155105.

CoMSES Net/OpenABM (2020), *Computational Model Library*. https://www.comses.net/.

Edmonds, B. and R. Meyer (2017), *Simulating Social Complexity*, *Simulating Social Complexity*, Berlin, Heidelberg: Springer. doi: 10.1007/978-3-540-93813-2.

Epstein, J.M. (2006), *Generative Social Science: Studies in Agent-Based Computational Modeling*, Princeton, NJ: Princeton University Press. doi: 10.5038/2162-4593.11.1.8.

Feliciani, T., A. Flache and J. Tolsma (2017), "How, when and where can spatial segregation induce opinion polarization? Two competing models", *Journal of Artificial Societies and Social Simulation*, **20** (2): 6. doi: 10.18564/jasss.3419.

Flache, A. (1996), *The Double Edge of Networks. An Analysis of the Effect of Informal Networks on Cooperation in Social Dilemmas*, Amsterdam: Thesis Publishers.

Flache, A. (2018), "About renegades and outgroup haters: Modeling the link between social influence and intergroup attitudes", *Advances in Complex Systems*, **21** (6–7): 1850017. doi: 10.1142/S0219525918500170.

Flache, A. (2019), "Social Integration in a Diverse Society: Social Complexity Models of the Link between Segregation and Opinion Polarization", in F. Abergel, B.K. Chakrabarti, A. Chakraborti, N. Deo and K. Sharma (eds), *New Perspectives and Challenges in Econophysics and Sociophysics* (*New Economic Windows*), Springer, Cham, pp. 213–28. doi: 10.1007/978-3-030-11364-3_15.

Flache, A., D.M. Bakker, M. Mäs and J. Dijkstra et al. (2017), "The Double Edge of Counter-Sanctions. Is Peer Sanctioning Robust to Counter-Punishment but Vulnerable to Counter-Reward?", in B. Jann and W. Przepiorka (eds), *Social Dilemmas, Institutions, and the Evolution of Cooperation*, Berlin: De Gruyter Oldenbourg, pp. 280–301. doi: 10.1515/9783110472974-014.

Flache, A., M. Mäs, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet and J. Lorenz (2017), "Models of social influence: Towards the next frontiers", *Journal of Artificial Societies and Social Simulation*, **20** (4): 2. doi: 10.18564/jasss.3521.

Flache, A. and M.W. Macy (1996), "The weakness of strong ties: Collective action failure in a highly cohesive group", *Journal of Mathematical Sociology*, **21** (1–2): 3–28. doi: 10.1080/0022250X.1996.9990172.

Flache, A. and M.W. Macy (2011), "Small worlds and cultural polarization", *Journal of Mathematical Sociology*, **35** (1–3): 146–76. doi: 10.1080/0022250X.2010.532261.

Fossett, M. (2006), "Including preference and social distance dynamics in multi-factor theories of segregation", *Journal of Mathematical Sociology*, **30** (3–4): 289–98. doi: 10.1080/00222500500544151.

Goldstein, D.G. (2009), "Heuristics", in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press, pp. 140–67. doi: 10.1093/oxfordhb/9780199215362.013.7.

Grimm, V., S.F. Railsback, C.E. Vincenot, U. Berger, C. Gallagher, D.L. DeAngelis, B. Edmonds, J. Ge, J. Giske, J. Groeneveld, A.S.A. Johnston, A. Milles, J. Nabe-Nielsen, J.G. Polhill, V. Radchuk, M.-S. Rohwäder, R.A. Stillman, J.C. Thiele and D.I. Ayllón (2020), "The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism", *Journal of Artificial Societies and Social Simulation*, **23** (2): 7. doi: 10.18564/jasss.4259.

Hedström, P. and P. Bearman (2009), "What Is Analytical Sociology All About? An Introductory Essay", in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press, pp. 3–24.

Hedström, P. and P. Ylikoski (2010), "Causal mechanisms in the social sciences", *Annual Review of Sociology*, **36** (1): 49–67. doi: 10.1146/annurev.soc.012809.102632.

Hegselmann, R. (2017), "Thomas C. Schelling and James M. Sakoda: The intellectual, technical, and social history of a model", *Journal of Artificial Societies and Social Simulation*, **20** (3): 15. doi: 10.18564/jasss.3511.

Hegselmann, R. and A. Flache (1998), "Understanding complex social dynamics: A plea for cellular automata based modelling", *Journal of Artificial Societies and Social Simulation*, **1** (3). http://www.soc.surrey.ac.uk/JASSS/1/3/1.html.

Holm, S., R. Lemm, O. Thees and L.M. Hilty (2016), "Enhancing agent-based models with discrete choice experiments", *Journal of Artificial Societies and Social Simulation*, **19** (3): 3. doi: 10.18564/jasss.3121.

Keijzer, M.A., M. Mäs and A. Flache (2018), "Communication in online social networks fosters cultural isolation", *Complexity*, 9502872. doi: 10.1155/2018/9502872.

Keuschnigg, M., N. Lovsjö and P. Hedström (2018), "Analytical sociology and computational social science", *Journal of Computational Social Science*, **1** (1): 3–14. doi: 10.1007/s42001-017-0006-5.

Langfred, C.W. (1998), "Is group cohesiveness a double-edged sword?", *Small Group Research*, **29** (1): 124–43. https://doi.org/10.1177/1046496498291005.

Macy, M.W. and A. Flache (2002), "Learning dynamics in social dilemmas", *Proceedings of the National Academy of Sciences of the United States of America*, **99** (suppl. 3): 7229–36. doi: 10.1073/pnas.092080099.

Macy, M.W. and A. Flache (2009), "Social Dynamics from the Bottom Up: Agent-based Models of Social Interaction", in P. Hedström and P. Bearman (eds), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press, pp. 245–68. doi: 10.1093/oxfordhb/9780199215362.013.11.

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

Macy, M.W. and M. Tsvetkova (2015), "The signal importance of noise", *Sociological Methods and Research*, **44** (2): 306–28. doi: 10.1177/0049124113508093.
Manzo, G. (2014), "Data, Generative Models, and Mechanisms: More on the Principles of Analytical Sociology", in G. Manzo (ed.), *Analytical Sociology*: *Actions and Networks*, Chichester, UK: Wiley pp. 4–52. doi: 10.1002/9781118762707.ch01.
Manzo, G. (2021), *Agent-based Models and Causal Inference*, Hoboken, NJ: Wiley.
Manzo, G., S. Gabbriellini, V. Roux and F.N. M'mbogori (2018), "Complex contagions and the diffusion of innovations: Evidence from a small-N study", *Journal of Archaeological Method and Theory*, **25** (4): 1109–54. doi: 10.1007/s10816-018-9393-z.
Mark, N.P., L. Smith-Lovin and C.L. Ridgeway (2009), "Why do nominal characteristics acquire status value? A minimal explanation for status construction", *American Journal of Sociology*, **115** (3): 832–62. doi: 10.1086/606142.
Mäs, M. (2021), "Analytical Sociology and Complexity Research", in G. Manzo (ed.), *Research Handbook on Analytical Sociology*, Cheltenham, UK: Edward Elgar.
Mäs, M. and A. Flache (2013), "Differentiation without distancing. Explaining bi-polarization of opinions without negative influence", *PLoS ONE*, **8** (11), p. e74516. doi: 10.1371/journal.pone.0074516.
Mäs, M., A. Flache, K. Takács and K.A. Jehn (2013), "In the short term we divide, in the long term we unite: Demographic crisscrossing and the effects of faultlines on subgroup polarization", *Organization Science*, **24** (3): 716–36. doi: 10.1287/orsc.1120.0767.
Mäs, M. and D. Helbing (2020), "Random deviations improve micro–macro predictions: An empirical test", *Sociological Methods and Research*, **49** (2): 387–417. doi: 10.1177/0049124117729708.
Mäs, M. and K.-D. Opp (2016), "When is ignorance bliss? Disclosing true information and cascades of norm violation in networks", *Social Networks*, **47**: 116–29. doi: 10.1016/j.socnet.2016.05.004.
Railsback, S.F. and V. Grimm (2019), *Agent-Based and Individual-Based Modeling: A Practical Introduction*, 2nd ed, Princeton, NJ: Princeton University Press.
Reutlinger, A., D. Hangleiter and S. Hartmann (2017), "Understanding (with) toy models", *The British Journal for the Philosophy of Science*, **69** (4): 1069–99. doi: 10.1093/bjps/axx005.
van de Rijt, A. (2019), "Self-correcting dynamics in social influence processes", *American Journal of Sociology*, **124** (5): 1468–95. doi: 10.1086/702899.
van de Rijt, A., D. Siegel and M.W. Macy (2009), "Commentary and Debate", *American Journal of Sociology*, **114** (4): 1166–80. doi: 10.1086/588795.
van Gent, W., M. Das and S. Musterd (2019), "Sociocultural, economic and ethnic homogeneity in residential mobility and spatial sorting among couples", *Environment and Planning A*, **51** (4): 891–912. doi: 10.1177/0308518X18823754.
Sakoda, J.M. (1971), "The checkerboard model of social interaction", *The Journal of Mathematical Sociology*, **1** (1): 119–32. doi: 10.1080/0022250X.1971.9989791.
Schelling, T.C. (1971), "Dynamic models of segregation", *The Journal of Mathematical Sociology*, **1** (2): 143–86. doi: 10.1080/0022250X.1971.9989794.
Simon, H.A. (1982), *Models of Bounded Rationality*, Cambridge, MA: MIT Press. doi: 10.7551/mitpress/4711.001.0001.
Squazzoni, F. (2012), *Agent-Based Computational Sociology*, Chichester, UK: Wiley. https://doi.org/10.1002/9781119954200.
Squazzoni, F., J.G. Polhill, B. Edmonds, P. Ahrweiler, P. Antosz, G. Scholz, É. Chappin, M. Borit, H. Verhagen, F. Giardini and N. Gilbert (2020), "Computational models that matter during a global pandemic outbreak: A call to action", *Journal of Artificial Societies and Social Simulation*, **23** (2): 10. doi: 10.18564/jasss.4298.
Wilensky, U. (1997), "NetLogo Segregation model", Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University. http://ccl.northwestern.edu/netlogo/models/Segregation.
Wilensky, U. (1999), "NetLogo". Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University. http://ccl.northwestern.edu/netlogo/.
Wilensky, U. and W. Rand (2015), *An Introduction to Agent-Based Modeling: Modeling Natural, Social and Engineered Complex Systems with NetLogo*, Cambridge, MA: MIT Press.
Wittek, R.P.M., T.A.B. Snijders and V. Nee (2013), "Introduction: Rational Choice Social Research", in *The Handbook of Rational Choice Social Research*, Stanford, CA: Stanford University Press, pp. 1–30.
Wooldridge, M. and N.R. Jennings (1995), "Intelligent agents: Theory and practice", *The Knowledge Engineering Review*, **10** (2): 115–52. doi: 10.1017/S0269888900008122.
Xie, Y. and X. Zhou (2012), "Modeling individual-level heterogeneity in racial residential segregation", *Proceedings of the National Academy of Sciences*, **109** (29): 11646–51. doi: 10.1073/PNAS.1202218109.
Young, H.P. (1998), *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Princeton, NJ: Princeton University Press.
Zhang, J. (2004), "A dynamic model of residential segregation", *Journal of Mathematical Sociology*, **28** (3): 147–70. doi: 10.1080/00222500490480202.
Zschache, J. (2018), "Melioration learning in iterated public goods games: The impact of exploratory noise", *The Journal of Mathematical Sociology*, **42** (1): 1–16. doi: 10.1080/0022250X.2017.1396983.

Andreas Flache and Carlos A. de Matos Fernandes - 9781789906851

25. The many model approach

Scott E. Page

Analytical sociology explains social facts such as residential segregation, cultural norms, and social mobility through mechanistic models that build from individual actions (Hedström and Bearman 2011). Those models can be expressed mathematically, computationally, or diagrammatically. To an external observer or a part-time consumer, the field's ever-growing collection of illustrative, thought-provoking, and intuition-reversing models reveals a wondrous complexity of aggregation processes and the necessity of engaging the micro-level complexity of the social world.

The contributions of the field derive in large part from the insights generated by these mechanistic unpackings of the linkages between micro- and meso-level behaviors and macro-level processes and patterns (Little 2012). These unpackings are necessary because aggregation, climbing the hull of Coleman's boat from micro to macro, creates no end of paradoxes: issue by issue majority rule can produce a majority-dominated policy (Ostrogorski's Paradox), an alternating combination of losing strategies can win (Parrondo's Paradox), and a trend can exist for each group, but disappear or reverse for the aggregate (Simpson's Paradox). Of course, sometimes the aggregation does add up: if everyone earns a higher income, total income increases, and if everyone loses weight, obesity declines. But more often than we might think, macro patterns need not align with micro behaviors, and in these cases, we need models.

In this chapter I argue that, in appraising the contributions of analytical sociology, we should look beyond the models' individual contributions and also contemplate their value in ensembles. The whole – the collection – may be more valuable than its parts in helping us to make sense of the world. To be more precise, if we examine any empirical macro-level pattern or trend – rising rates of obesity or opioid deaths, growing inequality, or the spread of a virus – multiple micro-level processes likely produce it. Thus, we benefit from multiple models, and the contribution of any one of them consists not just in its standalone explanatory power, but also in how it fits within the constellation of existing models. Extending that logic, the future models that may be most valuable may not be those that are most impressive in isolation, but those that build connective tissues with models that have already been constructed.

Just as a cathedral consists of more than a heap of stones joined together, a discipline consists of more than a random collection of ideas (von Mises 1949). And yet, constructing either a discipline or a cathedral involves placing one block atop and adjacent to others over long periods of time—the National Cathedral in Washington, DC took eighty-three years to complete, Notre Dame in Paris one hundred years, and the Sagrada Familia in Barcelona should accept its final brick in 2026, one hundred and forty years after construction began. The strength and promise of analytic sociology depends in part on how well its models sit adjacent and atop one another. In particular, do they constitute a collectively intelligent assemblage of models?

We can distinguish between two forms of collectively intelligence assemblages. First, an assemblage or ensemble can consist of a *variety* of models of the same type. For example, racial segregation or epidemiological models may make slightly different

Scott E. Page - 9781789906851

assumptions about social networks, behavioral rules, and adaptation. Variety ensembles create *improvements* in explanations, understandings, and prediction. These improvements can be measured quantitatively using R-squared measures of predictive accuracy. Or, the improvement can be qualitative, where the collection of models produces a more robust explanation or insight.

Second, an ensemble can be comprised of *diverse* models, that is, models of different types, with different assumptions. As discussed later in this chapter, economic inequality can be explained by models based on technology, human capital, residential sorting, or assortative mating (Page 2018). Each of these models assumes different primitives at the micro level – technology and labor markets, learning, risk taking, or partner selection. Well-constructed ensembles of diverse models produce an *emergent* collective intelligence – insights and understandings emerge that are not present, and not possibly present, in any of the parts. More, in these cases, can be different (Anderson 1972).

Whereas improving collective intelligence enriches our understanding of a particular phenomenon, emergent collective intelligence implies an entirely new understanding. To stretch a metaphor, individual neurons perform simple operations. Multiple neurons could perform similar operations and produce improved performance through error reduction. That would be improvement through an ensemble. A network of neurons can also perform much more elaborate functions. They can make calculations, feel emotion, and produce conscious thought and planning. These functions lie outside the repertoire of singular neurons. Similarly, cultures and other social regularities are also said to emerge (Sawyer 2005).

With ensembles of models, emergence occurs at the level of understanding. No single, comprehensible model can unpack all of the causes of obesity, economic or racial inequality, or social mobility. Diverse ensembles can. They can produce this comprehensive understanding, provided the constituent models shine light on distinct sets of attributes and processes, and overlap, by that I mean, they must make common assumptions and structures with other models.

Among the central points of this chapter will be that though any one model can help us to reason, explain, design, communicate, act, predict, or explore, ensembles of models preform these tasks better (Page 2018). Just as two heads often can be better than one, so too can two models, or better yet, three, or four, or five. I construct this argument in three parts. First, relying on a variation of Coleman's boat diagram, I formalize what I mean by improving and emergent collective intelligence. Second, I describe how a variety of models produces improving collective intelligence with an application to the spread of COVID-19. Third, using information theory, I introduce the *Borgian Hypercube* as a framework for understanding how a diversity of models can produce emergent collective intelligence. I then briefly describe the ensemble approach as it applies to rising income inequality.

1. MODEL VARIETY AND DIVERSITY

Social scientists primarily use models to reason through logic (to understand processes), to explain empirical phenomena, and to communicate ideas. Crisp logical models provide a counterpoint to richer, descriptive accounts that may require hundreds of pages. Over the past two decades, models have also become the standard methodology used in finance,

Scott E. Page - 9781789906851

politics, sports, and strategy. Predictions based on models took on special salience during the COVID-19 pandemic, in which epidemiological models that embed ideas from analytical sociology were used to predict to the virus's spread across communities and the effects of interventions such as lockdowns and contact tracing, a topic I return to in the next section.

To these four uses – to reason, explain, communicate, and predict – I add three more. Models can *inform* actions and organizational strategies, and then can aid in the *design* of better policies, programs, and interventions by helping us think through consequences. Finally, and this may be most understated if not under-appreciated value of models, they allow us to *explore* (Page 2018). With models we can run counterfactuals, and we can contemplate alternative worlds. A model need not be realistic to provide insights and understanding. One might take the position that many of the most influential computational models – Conway's Game of Life, Axelrod's Culture Model (1997), Watts and Strogatz's Small World Model (1998), and Schelling's Segregation Model (1978) have as much value as exploratory investigations in providing explanations of empirical phenomena.

While analytical sociologists primarily build mechanistic models to understand and explain social facts, many of their models find uses in guiding action, aiding design, communicating ideas, predicting outcomes, and exploring alternatives. Mechanistic models of social interactions can be represented graphically with *Coleman's Boat Diagram*; a slightly modified version is shown in (Figure 25.1) (Coleman 1990; Page 2016).

Analytical sociological models run below the surface, working along the bottom layer of Coleman's boat. Such models with a set of agents are represented in the bottom-left corner of the diagram. I will take the agents to be people, but they could also be organizations, algorithms, or nations. In agent-based models, the agents, arranged in a social structure (a network), follow behavioral rules encoded as computer algorithms. In mathematical models, the modeler derives the aggregate behavior, often an equilibrium. Arrow ① in the diagram represents how the micro-level behaviors aggregate to produce macro-level variables – levels of segregation, inequality, or obesity. Arrow ② along the bottom represents the micro-level interactions, from time period t to time period $t + 1$. The arrow on the right side of the diagram is also denoted with a ① because the aggregation process should be the same in each time period. This diagram extends Coleman's boat by including an arrow ④ connecting macro phenomena at time t with the micro-level behavior at time $t + 1$ to account for the possibility that individual behavioral rules depend on macro level variables. Some macro variables, such as the proportion of people infected with a disease, are mere algebraic aggregates of agent-level variable values.

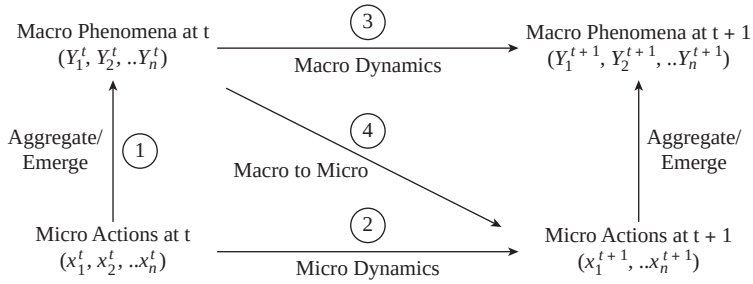


Figure 25.1 The aggregation diagram

Scott E. Page - 9781789906851

Others emerge; that is, they are not derived from mere algebraic calculations. In either case, micro-level behaviors may be more accurately described as functions of both their micro-level interactions and features of the macro environment.[1]

The Coleman's boat diagram implies a coherence in that the mapping across the top of the diagram agrees with the mapping across the bottom of the diagram mapped back to the top. A micro-level model might, therefore, produce aggregate patterns (represented by arrow ③) that can be described by an equation-based model riding along the gunwales that fits the relevant data. Thus, the best analytical sociology models might be said to "float Coleman's boat".

The behavioral rules noted are a key element in analytical sociology models. The rules may be habitual, culturally constrained, emergent, or the product of rational calculation. Note the departure from neoclassical economic rational choice models. Rational choice assumes calculating agents who anticipate outcomes if not with perfect accuracy, at least in expectation. Analytical sociology allows for other motivations. Herein lies the first opportunity, or might I say, necessity of model variety. We cannot know behavioral rules exactly. By constructing a variety of models with different behavioural assumptions, we can construct more robust inferences, make more accurate predictions, and reach deeper understandings, as discussed in the next section.

In addition, the agents who employ the behavioral rules are situated in social spaces – often represented as networks. Network configurations can have dramatic effects on outcomes. Interactions in highly connected networks or in sparse networks may produce less complex dynamics than interactions in moderately connected networks, what Miller and Page (2007) call "the interesting in between". Models of financial systems, simulations of the prisoners' dilemma game, and biological experiments all reveal complexity in regions of intermediate complexity (Golub and Jackson 2012; Kerr et al. 2002).

In computational models, the modeler presses the "go" button and analyzes what emerges (Miller and Page 2007; Epstein 2014). In addition to producing complexity, models can generate equilibria, regular cycles, or random sequences. All four classes of outcomes are seen in models and (to first approximation) in real-world data. Attendance at a weekly social event may reach a stochastic equilibrium, as shown in the El Farol Model (Arthur 1994). Stable cyclic patterns of work and leisure may emerge with clusters of individuals (Vagni and Cornwell 2018). Stock prices may be approximately random, while commodity prices complex. The complexity of a process can be defined as the difficulty of predicting accurately. Social outcomes—such as whether someone goes on welfare or graduates from college—are complex by this criterion, as shown in a recent competition involving hundreds of teams trying to make such predictions (Salganik et al. 2019).

An under-appreciated contribution of computational models in general, and analytical sociology models in particular, is that they draw our attention to the class of outcome: equilibrium, periodic, random, or complex. In mathematical models, especially rational choice models in which everyone anticipates what everyone else does, the outcomes are almost always equilibrium points or paths. Yet, when we look out at our social world, we see that many parts of it are not in equilibrium. Models help us to understand why.

---

[1] I drew diagram with only two levels. I could well include meso levels such as groups, communities, and organizations (Little 2012).

Scott E. Page - 9781789906851

2. MODEL COMPLEXITY

The Coleman's boat diagram draws a clean distinction between micro- and macro-level processes. It does not answer the question of how realistic or how granular to make the micro-level model. How many attributes define an agent? How elaborate and heterogenous are the behavioral rules? How empirically grounded are the assumptions about social networks? Increasing granularity and, therefore, complexity, offers more realism and potentially greater accuracy, but at a cost of transparency. In an often-cited essay titled "On Exactitude in Science", Borges describes a country that finds great value in maps. Eventually, their mapmakers construct a map of the country on a scale of a mile to a mile. The map never gets unfolded over objections from farmers who point out that unfolding the map would block out all sunlight and ruin their crops. Borges notes that instead of using the map, the people arrive at the ingenious idea to instead use the country as a map of itself.

This same logic can be applied to models, and this was Borges's point. One could construct the equivalent of Borges's mile-to-mile map, a model of precise exactitude. One reading of Borges is that such a model would be useless, no better than the world itself. That conclusion holds for some model uses, but not all. The amount of model complexity allowed depends on the use. To be transparent – to reveal mechanistic logic – a model must simplify, but for other uses, models might well be as complex as the world they capture and be of tremendous use.[2] Figure 25.2 graphically represents the relative maximum complexity of models for each of the seven categories of model uses.

By complexity, I mean the information content of the model – roughly speaking, how long it would take to explain or describe the model. For communication, a model must be relatively simple. At the other extreme, for prediction, an exact model would be ideal. For taking actions and designing, we might want a little less complexity because we would like to have some hope of understanding why we get the outcomes we do, though if we want full understanding, a model will have to be simple. Similarly, to reason, to understand the core mechanistic logics, a useful model would be quite limited in its complexity. Finally, when we use models to explore, we can allow for greater complexity as we need not understand fully how the model unfolds to acquire new intuitions. That said, we explore to learn about the implications of different assumptions, structure, and rules, so the model cannot be so complex as to be impenetrable.

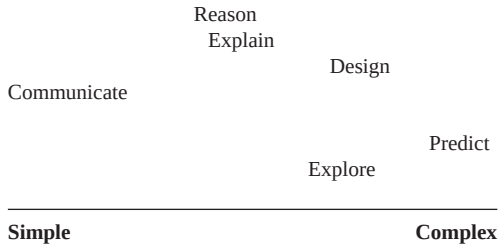


Figure 25.2 Maximal model complexity depends on use of model

[2] I thank Robert Axtell for pointing me to this line of reasoning.

Scott E. Page - 9781789906851

To summarize, Borges's mile-to-mile map model is not as silly as often made out to be, at least for some purposes. The mile-to-mile map, if we could construct it, would have real value as a policy tool. So too, can very simple models, and models whose complexity and granularity just exceed our powers of inference. Though it is not a central point of this essay, the fact that each of the seven uses of models recommends a different level of granularity provides another reason for many models. In any domain, we may seek to understand, predict, communicate, and explore. Each use might well involve a different model.

3. A VARIETY OF MODELS

I now turn to how multiple variants within a class of models that can lead to more accurate predictions, more complete explanations, and deeper understanding. I provide brief explanations of the predictive and explanatory benefits of ensembles given that I cover them in depth elsewhere (Page 2018). To start, assume we have N models that each makes a numerical prediction of the number of new cases of a virus in one month's time. Let M_i be the prediction of model i and let Θ be the true value that we learn in one month. Define the ensemble's prediction A to equal average of the N models' predictions. The *Diversity Prediction* states that the squared error of A equals the average squared error across the models minus the diversity of the models (the variance) (Page 2008):

$$(A-\Theta)^2 = \sum_{i=1}^{N}(M_i-\Theta)^2 - \sum_{i=1}^{N}(M_i - A)^2$$

Ensemble Error = Average Error – Diversity

The theorem explains why a variety of equally accurate models necessarily improves predictive accuracy. The squared error of the ensemble's prediction increases with the diversity (or variance) of the individual models' predictions. This occurs because holding the errors of each model constant, diversity correlates with the frequency with which models make errors in opposite directions. Two models might each miss the mark by six, but if one predicts too high and the other predicts too low, the average will be exactly correct. The crucial assumption is that model variety creates diverse predictions.

A similar logic holds for explanatory models. Here, we might use R-squared—the average squared error across a set of data—as a criterion. In the predictive context, we calculate the squared error of a single prediction. In the case of understanding, we calculate the squared error of a model applied to a collection of data. In this latter case, we decompose the squared error, the *model error*, into *categorization error* and *valuation error.* Categorization error results from the fact that the model partitions the data into categories. Any variation that remains within a category will be a part of model error. For example, in predicting income, if the model categorizes people by age, gender, education, and race, then all people who share the same vector of characteristics will be assigned the same expected income V_i. Valuation error refers to the difference between the value the model assigns to category i, denoted by M_i, and the true mean value for that category, V_i.

Given those definitions, any model satisfies the following equality:

$$\sum_{x\in X}\left(M(x)-V(x)\right)^2 = \sum_{i=1}^{n}\sum_{x\in S_i}\left(V(x)-V_i\right)^2 + \sum_{i=1}^{n}\left(M_i - V_i\right)^2$$

Scott E. Page - 9781789906851

This can be written in words as follows:

Model Error = Categorization Error + Valuation Error

The contributions of model variety are two-fold. It can increase R^2 by creating different categorizations or through different valuations. Either will produce variations in valuations. In each case, we need only refer back to the Diversity Prediction theorem to see how the average of a variety of models will provide a better explanation (as measured by squared error). Of course, the ensemble could contain more attributes, but no one model need include more. In other words, all of the models could be understandable, yet through their variety provide a better explanation of the variation in the data.

4. MODEL VARIETY IN THE SIR MODEL

I next provide a more detailed description of how model variety can improve performance by considering a variety of micro foundations for the SIR model from epidemiology.[3] The SIR model partitions a population into three categories: susceptible (S), infected (I), and recovered (R). Elaborations of the SIR model can create finer partitions that distinguish between asymptomatic and symptomatic people as well as the length of time someone has been infected. For clarity, I restrict attention to the simple form of SIR model described.

In discrete form, the number of newly infected people in a population of size N at time $t + 1$ equals the product of a contagion parameter c, the number of infected people at time t, and the proportion of susceptible people, $\frac{S_t}{N}$. The number of susceptible people decreases by that same amount. The number of recovered people, R_t, increases by a recovery rate a times number of infected people. These assumptions produce the following macro-level difference equation for the number of infected people at time $t + 1$.

$$I_{t+1} = cI_t \frac{S_t}{N} - aI_t$$

These equations connect macro-level variables over time and would be represented along the top of Coleman's boat. This model assumes only the most minimal micro foundations in that it is consistent with *random mixing* of individuals.

Early in the spread of a disease, the proportion of susceptible people approximately equals the population size. By inspection, the number of infected increases if and only if the contagion parameter c exceeds the recovery rate a. Epidemiologists refer to the ratio of these parameters $\frac{c}{a}$ as the *basic reproduction number*, R_0. Note that the basic reproduction number depends on behaviour. If R_0 exceeds one by a substantial margin, and if the disease causes discomfort, damage to health, or even death, then people will change their behaviors, which changes the value of c, the contagion parameter, thus changing the basic reproduction number.

In the case of the COVID-19 pandemic, people began social distancing, washing their hands more frequently, and, eventually, wearing face masks. Those adaptive responses reduced c, meaning that the basic reproduction number no longer captured the dynamics of the virus.

[3] This same model can be applied to the spread of technologies, ideas, customs, or laws.

Scott E. Page - 9781789906851

Hence, epidemiologists refer to the *effective reproduction number*, R_e. As data on the number of cases accumulates and behavior changes, epidemiologists update their estimates of R_e. During the COVID-19 pandemic any number of analysts estimate R_e for different regions. Those estimates informed policy actions. Government policies intended to slow or stop the spread of the disease did not, as some claimed, reduce R_0. That of course is impossible. R_0 is what it is. Instead, they reduced R_e, the effective reproduction number,

Early in the pandemic, the IHME, or Murray model, used to guide policy, relied on curve-fitting (Murray 2020). The original contained no micro foundations, not even a random mixing assumption. The model assumed a logistic function for the spread of COVID (a function for the macro-level process). Data were then used to estimate a logistic function. The modelers assumed that each of four interventions—school closings, nonessential business closings, stay-at-home orders, and travel restrictions—would shift the curve a third of the way to a lower curve. That curve, in turn, was estimated using data from regions in Italy and China that had imposed such measures. Eventually, the model was modified to include estimated SIR models, which are again largely lacking in micro foundations but led to a rather substantial re-estimation of number of fatalities. This can be interpreted as using a variety of macro-level models.

Simultaneously, epidemiologists, agent-based modelers, and analytical sociologists were working on a variety of models with micro foundations (Manzo and van de Rijt 2020; Wilder et al. 2020). These models assume individual agents embedded in networks characterized by multidimensional attributes (age, health status, employment status, race, etc.) and behavioral rules. The behavioral rules assumed vary depending on the model. The models differ in how they assume the model spreads, what social networks look like, and the extent of sequestration and by whom.

These micro foundations matter. The same SIR model could fit two very different diseases. For example, diphtheria and HIV each have estimated basic reproduction numbers of approximately three. Yet, at the micro level these two diseases could not be more different. HIV spreads through sexual contact, blood, and from mother to child. Each of these contact networks has a specific structure and each allows for better contact tracing than the spread of an airborne virus. People likely remember their sexual contacts, and women surely remember giving birth. In contrast, diphtheria spreads by saliva and droplets. Children who touch their mouths and each other more often than adults are particularly susceptible. Ironically, random mixing may not be a bad assumption for the social networks of children, particularly young children in a playground or in a preschool. These differences matter because of the potential for behavioral responses. Sexual contact network structures will change differently than children playgroup dynamics. For example, children lack agency, so the adults in the room can sever networks if they desire.

As this comparison makes evident, R_0 hardly embeds all the relevant attributes of a disease. Micro foundations can differ in *agent attributes*, *network structure*, and *behavioral rules*. Different choices in each of these three categories produces the model variety. Models could include any number of attributes: geographic location, age, gender, race, health status, income, wealth, profession, family size, and even political ideology. More attributes—provided you can calibrate them to data—could lead to more accurate predictions; however, include too many and the model may not be understandable.

Models can also differ in how elaborately they model networks. Some models calibrate to existing social network data (Manzo and van de Rijt 2020). This approach can help to explain regional variation through differences in social networks and demographics. A model built

Scott E. Page - 9781789906851

for Hubei may not be applicable to Lombardy, New York, or France (Wilder et al. 2020). That insight finds support in the enormous regional variation in the proportions of cases across and within countries. There's a catch though, in that those networks were estimated pre-COVID-19. What the modeler would like to know is how behavioral responses change the networks. For example, middle-aged people with young children and aged parents would be connectors who spread the disease from less cautious young people to older people if the networks did not change. Middle-aged people may limit their connections to one of the two groups. But how much? Different models will make different behavioral assumptions and produce different results. As shown in the two equations in the previous section, model variety will result in more accurate predictions and lower model error. Model variety also provides upper and lower bounds on what might occur. In the ensemble of models used by the IPCC to predict climate change, models vary in predicted average global temperature and change in sea levels. The minimum and maximum provide useful guidance for policy makers.

In addition, model variety improves our ability to reason. By testing varieties of a class of models we can learn what individual and joint changes will break a model, by that I mean produce qualitatively different outcomes. Consider the potential for a second wave of a COVID-19, or some other virus. We know from past epidemics, notably the 1918 flu, that second waves occur. They can be small and insignificant, or they can be large. One property we should require of a general purpose epidemic model would be that it can produce a large second wave (for some parameter values and behavioral assumptions), but that it does not always produce second wave (for others). Call this the *contingent second-wave property*. With any single-model variant, we can trace the macro-level patterns up to the current time that make a strong second wave probable, say greater than 50%. If we have a second, a third, and a fourth model, all of which satisfy the contingent second-wave property, we get alternative macro-level data patterns that suggest a strong second wave.

In Figure 25.3, each black circle represents possible data from the first wave of an epidemic. These might be the proportion of people infected in the first 30 days, the days until the first peak, and so on. The grey regions in top row show which first-wave data resulted in a second wave for three different models. The first model predicts second waves only if the first wave data fall in the grey square. The second model predicts a second wave for a wedge of first-wave data, and the third model predicts a second wave only if the first wave data falls within a small circle.

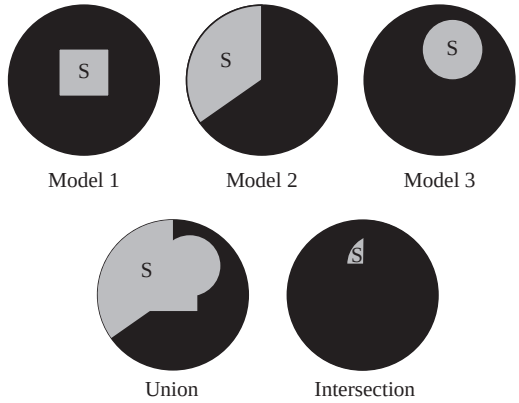


Figure 25.3 Three models and their conditions for second waves

Scott E. Page - 9781789906851

Each model identifies different sets of data to be consistent with a second wave. Those different sets provide opportunities for richer understandings. The circle in the bottom left shows the union of all three sets of conditions. If the first wave data falls in that region, then there exists at least one model consistent with a second wave. If the data lies outside this region, then the data are not consistent with a second wave in any model. The circle in the bottom right shows the intersection of the grey regions. If the first-wave data lie within that region, then the data are consistent with a second wave in every model.

Were we to increase model variety, we would increase the size of the set in the bottom left and decrease the set in the lower right. Suppose that we had a dozen models. Were we to encounter data that lies in the second-wave region for a single model or perhaps two, we could then interrogate those models in greater depth, exploring why they suggest second waves while others do not. The strong majority need not rule. We may find that these models include networks based on age and that other data show large disparities based on age. Or, we might find a common simplification in these two models, perhaps they assumed a skewed regional distribution of cases. We may find that micro-level assumption does not hold up to scrutiny, in which case, we would go with strong majority.

At the other extreme, the circle in the bottom right shows the intersection of the data sets consistent with a second wave for the three models. If the first-wave data fall within this set, then all models suggest a second wave. This does not guarantee a second wave – we must always keep in mind that all models are wrong. However, given that conditional on the first-wave data, three models with different micro foundations all predict a second wave, policy makers should take note. However, most policy makers would want to understand the proposed causes of the second wave, imposing a limit on how complex the models can be. Thus, the value of model variety as it enables the inclusion of multiple causal forces, if they were contained in a single model, might prove overwhelming.

I have described the choice of the variety of models as somewhat ad hoc. It might be history dependent, or driven by the whims or toolkits of the researchers. The nascent research program in *inverse generative social science* or *rule induction* relies on machine learning to automate the construction of a variety of models (Rand 2019). Given macro-level dynamics, the top of Coleman's boat, an algorithm searches through a variety of possible micro foundations to identify those that are consistent.

5. A DIVERSITY OF MODELS

Though the spread of a virus is a complex process, it is less complex than processes of central interest to sociology, such as the processes that generate economic and social inequality, health outcomes, segregation, and cultural change. The key difference is that the virus has a single cause – a pathogen. Inequality has multiple, interwoven causes – some economic, some social, and some behavioral. Understand these more complex social processes as a diversity of models: models that differ in their fundamental assumptions and structures.

To describe why understanding requires a diversity of models, I rely on concepts from information theory and construct a framework I call the *Borgian Hypercube*. I introduce both in the context of an investigation into the causes of rising obesity. The gains in obesity result from a variety of interacting causes. Figure 25.4 shows a simplified version of a much more complex spaghetti bowl of correlative and causal relations of the various contributors

Scott E. Page - 9781789906851

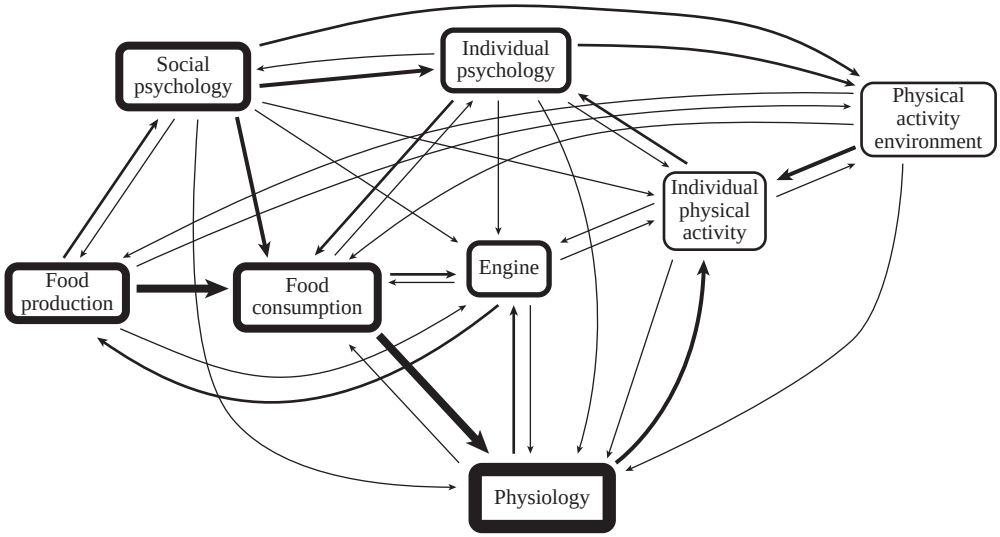


Figure 25.4 Simplified obesity causal map (Finegood et al. (2012))

to obesity (Finegood et al. 2012). The figure shows eight contributing systems showing that obesity is produced by a complex system of systems (Zellner and Page 2020).

For the purposes of elaboration, suppose that we possessed time series data for each box in this graph. With these data, we could compute the entropy, or information content, of each of the right contributing systems. For example, H_1 could denote the entropy of the biological data, H_2 could denote the entropy of the food data, and so on. We can let H_S denote the entropy of the entire system of systems. We can think of entropy as a proxy for the overall complexity with some caveats (Prokopenko et al. 2009). Given that the entire system consists of the eight systems, the following weak inequality holds:

$$H_S \leq \sum_{i=1}^{8} H_i$$

Assuming correlations across systems – the food, economics, and infrastructure systems are surely correlated – the inequality will be strict.

If we tried to create a model of the entire system, it might be of use as a predictor but have limited explanatory value as it is too complex. A many-model approach would construct separate models for each of the eight subsystems. Each model would produce a time series of data, which we could measure with entropy. The complexity of each model could also be measured using *minimal description length*. For the purposes of this article, I will assume that we have agreed on on complexity measure for each model and let $H_{M(i)}$ denote the complexity of the model of subsystem *i*.

The subsystem models are diverse. A model of food production has few links to models of the psychological processes that drive diversity. Also, to be understandable, each model must have limited complexity. I will assume a bound of $\bar{H}$. Any model more complex than $\bar{H}$ cannot be understood.

Scott E. Page - 9781789906851

This information theoretic lens implies Box's dictum that all models are wrong. If the $\bar{H}$ complexity constraint binds, as it surely will, then the model is less complex than the process it describes. Hence, the model must be wrong.

5.1 The Borgian Hypercube

The eight diverse models form the basis for the *Borgian Hypercube*, which consists of all possible combinations of the eight models. Each corner (vertex) on the Borgian Hypercube consists of a subset of the models. The corner (0,1,0,0,0,0,0,0) consists only of model two. The corner (1,1,0,1,0,0,0,0) consists of models one, two, and four. The corner (1,1,1,1,1,1,1,1) includes all eight models. It is not quite the mile-to-mile map, but it is not far from it. The Borgian Hypercube consists of 8 single models, 28 pairs of models, and 56 triples of models. There exist 255 model combinations plus a null model. No one person could explore all of these combinations, but a good-sized team of scholars could work through the combinations of three models or fewer.

At each vertex of the Borgian Hypercube, we can compute the information content or, if we were motivated by explanatory power, the adjusted R-squared, or some measure of how well the model predicts out of sample. Regardless of the summary statistic computed at each vertex, we should expect *subadditivity*: the information content, adjusted R-squared, or predictive power of the node that includes models 1, 2, and 3, should be less than the sum of that summary statistic for the models of systems 1, 2, and 3 considered in isolation. That subadditivity means that the models overlap in what they produce, explain, or predict.

The Borgian Hypercube contrasts with the oft-repeated parable of the blind men and the elephant. In the parable, the elephant consists of five attributes: a trunk, ear, leg, side, and tail. Each person experiences and interprets a single attribute. No one is aware of any connections between them. In the Borgian Hypercube, a person looking at any one model knows that it addresses obesity and, when that person adds a second model, it sees the connective tissue because the models overlap. Thus, the blind man touching the trunk, if you will, also rubs his hand along the side, as does the blind man touching the tail. Understanding one part of the system is not as misleading as the parable suggests.

As the number of component models grow, as we move toward the mile-to-mile map, the combined model becomes less interpretable. The process of understanding may, therefore, start with a single model and add and subtract other models, with the total number of combined models exceeding three or four. To be clear, I am not advocating step-wise regression to find the optimal collection of models or proposing as analog of the Lasso method from statistics. (Lasso, an acronym for least absolute shrinkage and selection operator, imposes a cost function on larger coefficients preventing overfitting.) My objective, and that of analytical sociology, is to compose a collection of models that best help us understand phenomena. By maneuvering around the hypercube, we learn how models interact. Some of the most useful models, to put an ironic twist on Box's famous dictum, may be those that are most wrong.

6. MANY MODELS OF INEQUALITY

I now briefly describe how one might apply the Borgian Hypercube framework to increase our understanding of income inequality. In Page (2018), I describe more than ten models that have been advanced to why income inequality has increased over the past quarter century in

Scott E. Page - 9781789906851

the United States and worldwide. Here, I focus on a smaller number of sociological models. I begin with some basic facts: most city-, regional-, and country-level income distributions can be calibrated to lognormal distribution or Pareto distributions. That means that they have elongated tails. When demographers and economists say that income inequality increases, they mean a stretching of the distribution, so that that the share of income going to the top 20%, 10%, 1% and 0.1% increases.

Neoclassical economic models explain increasing inequality as resulting from changes in the returns to educated and uneducated labor. The relative return to education changes over time, owing to technological change (Goldin and Katz 2008). These models divide people into categories based on education level. The equilibrium wage paid to a type of labor depends on the relative supply of that type and on current technology. Inequality results from changes in relative supply and demand. In the 1970s college graduates earned only a little more than high school graduates because of demand for physical labor. Automation and the digital age shifted labor demands so that now college graduates earn much more.

Economists also explain increasing inequality as resulting from rents and to excess returns to entrepreneurs. Piketty's bestselling book *Capital in the Twenty-First Century* describes how a standard neoclassical economic model produces returns on capital that exceed the growth rate in the economy ($r > g$). Thus, income from capital rents increases faster than income. Entrepreneurial models assume people's returns are proportional to their human capital. These models can explain the lengthening upper tail of the distribution: entrepreneurs make up nearly three-fourths of the 400 wealthiest individuals in the United States (Kaplan and Rauh 2013; Jones and Kim 2018). Such models can also explain why we see inequality within each profession – some people are just better than others. The best chefs, airline pilots, cellists, and personal trainers make enormous salaries (Frank 1996; Xie, Killewald and Near 2016).

We are now off in an economic corner of the inequality Borgian Hypercube. We might also start from more sociological foundations. In a rational choice framing, individuals choose optimal levels of education. In a sociology model, people may make choices similar to their friends, family, and fellow community members. The old saying that economics studies why people make choices and sociology studies why people lack agency to make choices, and this rings partly true. Analytical sociological models would also take into account physical location. Such models might include average income, levels of school funding, average educational attainment, crime rates, lead levels in water, and other health-related variables. When estimated or calibrated, these models reveal the impact of location on educational choices and, therefore, income. Empirical studies of the United States going back more than a half a century reveal substantial correlations. Poorer neighborhoods have higher crime rates, higher levels of lead, and less educational attainment. Of particular importance, these detrimental characteristics correlate to race. African Americans disproportionately live in neighborhoods with characteristics that negatively correlate with economic success.

Adding characteristics of place changes the economic returns from getting an education as some school systems offer fewer opportunities. In 2015–2016 half of Detroit Public Schools offered no advanced placement classes in any field. It also changes the culture within which educational choices are made. Both effects would result in fewer people in poor neighborhoods choosing to build human capital through education. In addition, in the educational context, the choices include not just level of education, but also areas of study, as well as specific tools and skills. A degree in machine learning or data science results in more lucrative job offers than a degree in sociology or English literature. A student who does not know any data scientists,

Scott E. Page - 9781789906851

who attends a school that does not offer computer science courses, much less advancement placement courses, would probably not be aware of the many career paths that knowledge of data science offers.

Correlations in disparities by location therefore contribute to inequality and its persistence. Sociologists refer to that correlation as *consolidation*, meaning that we need only one variable, race, to know a great deal about other variables of interest (Blau and Schwartz 1984). Public health scholars refer to the combination of causal forces as a *syndemic* (Singer and Clair 2003). Following Sampson (2019), I will refer to this constellation of negative causal attributes as *compounded disadvantage*. The analytical sociology models that produce compound disadvantage explain why some people fail to acquire the skills that lead to higher incomes as well as why generations within the same families become stuck in the lower end of income distribution. They lack of income mobility.

If we now add in an economic model, say the entrepreneurial model, we find unexpected connections. People who live in high-income communities will be connected indirectly to more people with access to economically valuable information. People in low-income neighborhoods will not. Data show, for example, racial differences in uptake of the internet (DiMaggio and Garip 2011). Data also show enormous racial differences in the likelihood of filing a patent or being an entrepreneur accounting for test scores (Bell et al. 2019). To connect the dots, compounded deprivation means less return to education; it also means less access to ideas and, therefore, less opportunity to be an entrepreneur. The entrepreneur model reveals an even greater compound disadvantage.

To add just one more model, *assortative mating* has been shown to explain a substantial portion of increasing inequality. This model is almost too simple: a family's income equals the sum of the incomes of the partners. If a low-income person marries a high-income person, then that marriage will contribute toward equalizing income distributions. If high-income people marry other high earners, then income disparities increase. If we add this model to the technology and labor model from economics, we see how increasing returns to education exacerbate inequality.

Women now earn college degrees at a higher rate than men. As noted, income is rising with education level. Educated women strongly prefer equally or more educated partners. This creates strong correlation in partners' incomes (Cancian and Reed 1999; Schwartz and Mare 2005). We can also add in a segregation model. If people live in neighorhoods segregated by income, and if people meet people locally or use locations learned on line as signals of status, then residential sorting will reinforce assortative mating. These effects can be substantial. Had marriages been random rather than assortative, income inequality would be much less. One study finds that inequality as measured by the *Gini Coefficient*, the distance between the income distribution and an equal distribution. By one estimate, the Gini coefficient would have decreased by 25% (see Greenwood et al. 2014).

The central takeaway is that the diverse models highlight different processes. Technological change models show a skew in the returns to education. Segregation and compound disadvantage models reveal less return to education and, when linked to social diffusion of information models, less chance of being an entrepreneur. Entrepreneurship models, in turn, explain how weightless goods create a lengthening tail. And, finally, assortative mating models show that if people prefer partners of similar status, families get two incomes from the upper tail, resulting in even more inequality. By adding and subtracting models, and moving about the Borgian Hypercube, we see the connective tissue, the full elephant, but we maintain understandability. We are not wandering about on the mile-to-mile map.

Scott E. Page - 9781789906851

7. SUMMARY

In this chapter, I put forth two separate arguments for using multiple models. First, I claim that a variety of models within a type leads to better predictions, explanations, and understandings. And, second, I claim that for complex systems composed of multiple processes, a diversity of models, that is, models of each process, enable a richer understanding of complex processes. The second claim builds from the concept of an ontology of a system as being a representation of actors or entities and their interactions. A careful ontology of any social process or institution reveals elaborate depth and complexity, as shown by Little (2020), who provides a rich ontological unpacking of government as a composite of functions, networks, and power dynamics operating on multiple levels. In the terminology I use here, the government is a complex adaptive system of systems. Multiple actors work in coordination and in competition with one another in a variety of capacities. A mile-to-mile map of government would not be understandable. Hence, we need many (diverse) models if we hope to achieve understanding.

I have suggested that the discipline of analytical sociology would benefit from encouraging connective models, that is, models that build on and relate to other models. The small worlds network model and the complex contagion model (Centola and Macy 2007) are two foundational bricks for a cathedral. Each can be plugged into many different models. Each also suggests many possible variations. Many models can be built from and connect to these models.

In constructing social science models, and especially analytical sociology models, we aim for a target with higher dimensionality and greater complexity than the frameworks we can construct. As we have learned, we cannot, as Plato suggests, carve nature at its joints. To do so carries the risk of missing the elephant in the room. Instead, we must explore the joints and connective tissues by wandering about on Borgian Hypercubes. By connecting models to other models, we may not reach full understanding, fully explain the world, or predict with perfect accuracy. We may find, ex post, that we have advocated suboptimal actions and kludgy designs. But we will do better than we would have done with individual models, and far better than if we had used no models at all. An apt aspirational metaphor might be Coleman's self-organizing fleet, heading out not in fixed formation, but schooling and exploring so as to meet the challenges ahead.

REFERENCES

Anderson, P. (1972), "More is different", *Science*, **177**: 393–6.

Arthur, W.B. (1994), "Inductive reasoning and bounded rationality", *American Economic Review Papers and Proceedings*, **84**: 406–11.

Axelrod, Robert (1997), "The dissemination of culture: A model with local convergence and global polarization", *Journal of Conflict Resolution*, **41**: 203–26.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, John Van Reenen (2019), Who becomes an inventor in America? The importance of exposure to innovation, *The Quarterly Journal of Economics*, **134** (2): 647–713.

Blau, Peter Michael and Joseph E. Schwartz (1984), *Crosscutting Social Circles: Testing A Macrostructural Theory of Intergroup Relations*, Orlando, FL: Academic Press.

Cancian, M. and D. Reed (1999), "The impact of wives' earnings on income inequality: Issues and estimates", *Demography*, **36**: 173–84.

Centola, Damon and Michael Macy (2007), "Complex contagions and the weakness of long ties", *American Journal of Sociology*, **113** (3): 702–34.

Coleman, James S. (1990), *Foundations of Social Theory*, Cambridge, MA: Belknap Press.

DiMaggio, Paul and Filiz Garip (2011), "How network externalities can exacerbate intergroup inequality", *American Journal of Sociology*, **116** (6): 1887–933.

Scott E. Page - 9781789906851

Finegood, Diane, Thomas D.N. Merth and Harry Rutter (2012), "Implications of the foresight obesity system map for solutions to childhood obesity", *Obesity*, S13–S16.
Frank, Robert (1996), *The Winner-Take-All Society: Why the Few at the Top Get So Much More Than the Rest of Us*, New York: Penguin.
Goldin, Claudia and Lawrence F. Katz (2008), *The Race Between Education and Technology*, Cambridge, MA: Harvard University Press.
Golub, Ben and Matthew Jackson (2012), "How homophily affects the speed of learning and best response dynamics", *Quarterly Journal of Economics*, **127** (3): 1287–338.
Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov and Cezar Santos (2014), "Marry your like: Assortative mating and income inequality." *American Economic Review: Papers & Proceedings*, **104** (5): 348–53.
Hedström, Peter and Peter Bearman (eds) (2011), "What is analytical sociology all about? An introductory essay", *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press.
Jones, Charles and Jihee Kim (2018), "A Schumpeterian model of top income inequality'', *Journal of Political Economy*, **126** (5): 1785–826.
Kaplan, Stephen and Joshua Rauh (2013), "It's the market: The broad-based rise in the return to top talent", *Journal of Economics Perspectives*, **27** (3): 35–56.
Kerr, B., M. Riley, M. Feldman and B. Bohannon (2002), "Local dispersal and interaction promote coexistence in a real life game of rock-paper-scissors", *Nature*, **418**: 171–4.
Little, Daniel (2012), "Explanatory autonomy and Coleman's boat", *Theoria: Revista de Teoría, Historia y Fundamentos de la Ciencia*, **27** (2): 137–51.
Little, Daniel (2020), *A New Social Ontology of Government Consent, Coordination, and Authority*, New York: Palgrave McMillan.
Manzo, Gianluca and Arnout van de Rijt (2020), "Halting SARS-CoV-2 by targeting high-contact Individuals", *Journal of Artificial Societies and Social Simulation*, **23** (4): 10. http://jasss.soc.surrey.ac.uk/23/4/10.html. doi: 10.18564/jasss.4435.
Miller, John and Scott E. Page (2007), *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*, Princeton, NJ: Princeton University Press.
Murray, Christopher and IHME COVID-19 Health Service Utilization Forecasting Team (2020), "Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months", medRxiv 2020.03.27.20043752.
Page, Scott E. (2016), "What should sociologists know about complexity", *Annual Review of Sociology*, **41**: 21–41.
Page, Scott E. (2018), *The Model Thinker*, New York: Basic Books.
Prokopenko, Mikhail, Fabio Boshectti and Alex J. Ryan (2009), "An information-theoretic primer on complexity, self-organization, and emergence", *Complexity*, **15** (1): 11–28.
Rand, William (2019), "Theory-Interpretable, Data-Driven Agent-Based Modeling", in P.K. Davis, A. O'Mahony and J. Pfautz (eds), *Social-Behavioral Modeling for Complex Systems*. doi:10.1002/9781119485001.
Salganik, Matthew J., Peter Dodds and Duncan Watts (2006), "Experimental study of inequality and unpredictability in an artificial cultural market", *Science*, **311**: 854–6.
Salganik, Matthew, Ian Lundberg, Alexander T. Kindel and Sara McLanahan (2019), "Introduction to the special collection on the fragile families challenge", *Socius*, **5**, Abstract.
Sampson, R.J. (2019), "Neighbourhood effects and beyond: Explaining the paradoxes of inequality in the changing American metropolis", *Urban Studies*, **56** (1): 3–32.
Sawyer, D. Keith (2005), *Social Emergence: Societies as a Complex System*, Cambridge, UK: Cambridge University Press.
Schelling, Thomas (1978), *Micromotives and Macrobehavior*, New York: Norton.
Schwartz, C.R. and R.D. Mare (2005), "Trends in educational assortative marriage from 1940 to 2003", *Demography*, **42**: 621–46.
Singer, Merrill and Scott Clair (2003), "Syndemics and public health: Reconceptualizing disease in bio-social context", *Medical Anthropology Quarterly*, **17** (4): 423–41.
Vagni, Giacomo and Benjamin Cornwell (2018), "Patterns of everyday activities across social contexts", *Proceedings of the National Academy of Sciences*, **115** (24): 6183–8.
von Mises, Ludwig (1949), *Human Action: A Treatise on Economics*, New Haven, CT: Yale University Press.
Watts, Duncan, J. and Steven H. Strogatz (1998), "Collective dynamics of 'small-world' network", *Nature*, **393** (6684): 440–2.
Wilder, Bryan et al. (2020), "Modeling Between-Population Variation in COVID-19 Dynamics in Hubei, Lombardy, and New York City", SSRN Posted: 31 Mar 2020. Last revised: 29 May 2020.
Xie, Yu, Alexandra Killewald, and Christopher Near (2016), "Between- and within-occupation inequality: The case of high status professions.'' *Annals of the American Academy of Political and Social Science*, **663** (1): 53–79.

Scott E. Page - 9781789906851

Coda—The past and future of analytical sociology[1]

Peter Hedström

1. INTRODUCTION

Twenty-five years now have passed since the Royal Swedish Academy of Sciences conference took place in Stockholm, leading to *Social Mechanisms: An Analytical Approach to Social Theory* (Hedström and Swedberg 1998), the book that, according to many, marked the beginning of analytical sociology as we perceive it today. This calls for some reflections on how the field has developed. Since the field has grown considerably during these years, my reflections will only cover a small subset of the issues that analytical sociologists are concerned with, and my set of priorities for the future may not necessarily be shared by everyone.

The chapter is organized as follows: First, I briefly discuss the birth of analytical sociology, what motivated its emergence, and what its most essential characteristics are. Second, and in more detail, I consider a set of topics that I believe to be of crucial importance for the future of analytical sociology. I pay particular attention to the role of action-based explanations of macro-level change. While the link between micro behaviors and macro dynamics is at the very core of analytical sociology, the importance many analytical sociologists attach to intentional explanations of behavior, combined with the lack of information on the individuals' intentions, has led to a proliferation of just-so stories. This sort of storytelling, based on the analyst's own preferred action theory, contradicts one of the main tenets of analytical sociology—that explanations should refer to the causal processes known to be responsible for the outcomes to be explained—and I suggest that we should avoid intentional explanations in all but the rare cases where we have access to reliable information on the mental states of the acting individuals. Instead we should concentrate our explanatory efforts on interaction structures and the macro patterns they give rise to.

2. THE CORE OF ANALYTICAL SOCIOLOGY

Since I already have written extensively on the core characteristics of analytical sociology (e.g. Hedström and Bearman 2009), I will keep this section brief. When writing *Dissecting the Social* (Hedström 2005), my ambition was to start from first principles: to identify what characterizes an ideal explanation, and then, from this ideal, derive how one ought to go about explaining social phenomena and how one ought to design empirical research in order for it to be as relevant as possible for the explanatory endeavor.

Sociological theory and research of the day had an ambivalent and poorly articulated relationship to explanations, particularly to the kind of mechanism- and process-oriented explanations that came to be adopted by analytical sociologists. Social theory was more oriented toward

[1] I would like to thank Marc Keuschnigg and Gianluca Manzo for their excellent comments on an earlier draft. The research has been supported by a grant from the Swedish Research Council (DNR 445-2013-7681).

Peter Hedström - 9781789906851

normative critique and conceptual elaborations than toward the development of explanatory theories, and empirical research was more concerned with describing social developments than to test and empirically identify the cogs and wheels of important social processes. Quantitative empirical research, by and large, was dominated by a rather superficial variable-centered form of analysis where, to use Abbott's (1998) apt expression, variables and not actors did the acting (see also Goldthorpe 2021 for an interesting discussion of this state of affairs).

The explanatory approach advocated by many analytical sociologists was a form of a mechanism approach with the following core characteristics (see also Manzo 2014):

1. The *typical explanandum* is some sort of macro-outcome or pattern. Macro outcomes can be observed at different scales, ranging from the small group to society at large, and outcomes of interest include traditional sociological concerns such as inequalities, segregation, and diffusion patterns, as well as, say, the growth dynamics of organizational populations or the structural properties of social networks.
2. An acceptable explanation of such outcomes details the *social processes* through which the outcome was brought about. Simply referring to a particular event as the cause of the outcome is not sufficient; the process through which the initial event brings about the outcome must also be specified.[2]
3. Individuals are the *basic entities* of these kinds of processes because, although variously constrained, it is their behavior that drives the processes forward.
4. Existing *relationships* between the individuals also are important because they often influence the individuals' behavior and thereby also the ways in which the social process unfolds.
5. The kind of explanation we seek takes the form of a clear, precise, and empirically supported *model* of a social process, which is such that it steers the group of individuals into eventually bringing about the type of outcome to be explained.

The importance of using model-based explanations derives from the fact that micro-macro dynamics tend to be highly complex. Particularly when many individuals are involved in them, it often is impossible to know whether a proposed explanation is at all relevant in the sense of being able to bring about the type of macro-outcome to be explained.

Because of its flexibility and its ability to also represent complex social processes, agent-based simulation modelling emerged as a central tool for analytical sociologists. It is a simple-to-use formalism that allows researchers to simulate various processes and thereby examine whether a proposed mechanism or process can generate the type of outcome to be explained (see Flache and de Matos Fernandes, this volume). Although many applications of agent-based modelling leave a lot to be desired from an explanatory point of view, the methodology as such is of crucial importance for analytical sociology (Hedström and Manzo 2015).

---

[2] In some of my previous publications I have put more emphasis on social *mechanisms* than on social *processes*. This change in emphasis, in part, reflects changed priorities. An explanation that refers to a mechanism without spelling out its implied dynamics is not of much use for explaining a macro outcome, while information on a social process, even when poorly articulated and understood, can provide important insights useful for our explanatory efforts. More generally, though, whether we emphasize one rather than another is rather inconsequential because, as Manzo (2007: 5–6) expressed it, a "process is nothing more than the dynamic aspect of one (or several) mechanism(s): it is what the mechanism can trigger".

Peter Hedström - 9781789906851

The type of model I have in mind here when I refer to model-based explanations also is some sort of agent-based simulation model, but clearly other types of models could be envisioned such as purely mathematical models or even exclusively verbal models. But for a verbal account to count as a model, in this respect, it must be so clearly and precisely stated that it directly and without any ambiguities can be translated into an agent-based simulation model. However, a serious drawback with purely verbal models is that they are not well-equipped for getting a handle on complex processes involving large numbers of interacting individuals.

In order to make these abstract explanatory principles more concrete, let us use as an example that currently is of utmost concern throughout the world, the COVID-19 pandemic. If we were to explain why the pandemic developed as it did in a certain area, a simplistic causal-effect approach relating some attributes of the area in question to the state of the pandemic would not suffice. Instead, we would have to follow the scheme outlined and clearly specify the core ingredients of the dynamic *process* through which the virus spreads in the population. Although the virus as such is the key to the epidemic, how the epidemic develops over time to a large extent is the result of a social process. The virus would have died in its infancy had it not been for individuals interacting with one another, and the size of the epidemic and its development over time is, to a large extent, driven by the interaction patterns between the individuals.

In order to explain why the epidemic developed as it did, we need to specify an empirically supported model of the process believed to have been at work. This model must represent the relevant group of individuals, the patterns of interaction between them, the risk that an infected individual transmits the virus to a susceptible individual, recovery rates, and so on. Once a model is at hand, we can examine its explanatory relevance by comparing the macro outcomes it generates with those observed in reality. If it passes this relevance test, we can use the model to address more specific queries about what is likely to have happened to the epidemic had the entities, relations, or activities that govern the process been subject to some sort of surgical intervention that did not affect any other part of the causal structure, e.g. interventions affecting the interaction patterns between the individuals such as travel restrictions or isolating or vaccinating high-contact individuals (for an interesting analysis of this latter point, see Manzo and van de Rijt 2020).

The explanatory strategy would have been exactly the same had we instead sought to explain, say, the ethnic segregation of a city, the emergence of status hierarchies, or downloads at Spotify. That is, we must specify an empirically supported model of the process through which a group of interacting individuals gradually bring about the outcome to be explained. However, as Centola has called attention to with his distinction between simple and complex contagion (see Centola, this volume), the process through which a behavior or an idea spreads from one individual to another often is more complex than the processes through which a virus spreads, but the explanatory principles are the same (see also Goldberg and Stein 2018).

What this explanatory scheme calls attention to is the importance of basing explanations of macro outcomes on empirically plausible and process-oriented micro foundations. Without such foundations we are likely to misconceive the complex causal structure of the social world, and we are likely to offer incorrect or at least vague and shallow explanations for why we observe what we observe. My sense is that the importance of using this approach has become gradually more recognized within the discipline (e.g. Bearman et al. 2004; Bruch and Mare 2006; Goldberg and Stein 2018). This positive development is not only due to the examples set

Peter Hedström - 9781789906851

by analytical sociologists, but also reflects the increased availability of data on large populations of interacting individuals and new methods suitable for analyzing such data (see Spaiser, this volume, Jarvis et al. (forthcoming), Keuschnigg et al. 2018).

3. THEORETICAL DEVELOPMENTS

The social mechanism approach, as originally perceived, not only called for empirically plausible and process-oriented models; it was also a call for a different kind of sociological theory. It insisted that we should abandon the idea of general theory, be it of the Parsonian kind of the past or the rational choice kind of the more recent past, and instead adopt a view of sociological theory as a toolbox of social mechanisms.

This vision of theory does not require that the theories in the toolbox eventually be organized into a grand unified theory. Nor does it require that the theories in the toolbox are true or false in any absolute sense of the word (Hedström and Ylikoski 2010). Instead, these theories have the character of what Coleman (1964) referred to as "sometimes-true theories". A theory of this kind, which details the logic of a certain type of social process, should not be evaluated in terms of whether it is true or false in general, but whether it applies to the specific situation at hand. In many instances several different kinds of processes could have brought about the outcome to be explained, and "the problem in any given application is to know *which* of the many processes is followed in this case" (Coleman 1964: 516–7) because the explanation must be done with reference to the type of process that indeed was at work in the empirical case to be explained.

This type of theory has much in common with Merton's (1967) idea of sociological theories of the middle range, something that Merton himself also strongly emphasized at the 1996 mechanism conference in Stockholm. Merton originally introduced the idea of middle-range theory in a comment on a paper by Parsons. In this comment he objected to Parsons's belief in the possibility of creating a sociological *theory*, in the singular, and the corollary belief that it would be possible to develop a theory that would be general enough to treat *all phenomena* of interest to sociology. "Sociology will advance in the degree that the major concern is with developing theories adequate to limited ranges of phenomena and it will be hampered if attention is centered on theory in the large" (Merton 1948: 165–6).

As discussed in Hedström and Udehn (2009), a theory of the middle range is a type of theory that clearly and precisely specifies a type of process that can bring about a particular type of outcome. The prototypical example is Merton's (1948) theory of the self-fulfilling prophecy, which details a process through which individuals, in interaction with one another, gradually make an initially false belief about the world come true. The theory is general in the sense that it applies to many different types of beliefs and behaviors, but the explanation is partial in the sense that it does not seek to explain token outcomes in all their minute details. This type of theory instead seeks to bring out the heart of the matter by detailing a *type* of process that can explain the *type* of outcome that the token outcome is an example of.

While there has been considerable methodological progress in how sociologists go about analyzing and explaining macro-outcomes, the theoretical toolbox has not progressed in the same manner. At the close of the 1996 Stockholm conference on social mechanisms, Thomas Schelling suggested that the next important step for us should be to systematize sociological theory by creating better order in the theoretical toolbox. To do this, he suggested, we should

Peter Hedström - 9781789906851

initiate a book project that arranged and categorized the already known social mechanisms and identified those areas of importance where the existing toolbox was shallow or empty. Unfortunately, that never happened, and the lack of progress in this area is perhaps most clearly seen in that we still use Merton's old theories to illustrate what kind of mechanism theories we have in mind. With a risk of sounding unappreciative, which I certainly am not, although the very best examples of analytical sociology, such as the winners of the Robert K. Merton best article awards, clearly articulate the type of processes that have brought about the outcomes they seek to explain (see Manzo, this volume), they have not produced *theories* of the same compelling simplicity and elegance as Merton's.

It is perhaps to demand too much of a field that it should produce theories on par with those of Merton. After all, the entire history of sociology has not produced all that many of them. I do believe, however, that a more rapid theoretical progress would have been possible, but that it has been held back by the disciplinary culture in which we are embedded. In this culture, as emphasized by Kroneberg and Tutic (this volume), theorizing tends to be equated with proposing concepts, heuristics, and frameworks, and theoretical progress with the development of increasingly differentiated and complex conceptualizations of actions and practices. This idea of what theory and theorizing is all about stands in sharp contrast to the analytical-sociology view of *theory as explanation*, and to the importance it attaches to clear and precise specifications of different types of social processes.

Being embedded in a somewhat foreign theoretical culture has many unfortunate consequences. One is that those who have the analytical ability to fully understand the complexities involved in these types of social processes typically do not view themselves as theorists and do not see it as their role to contribute to the theoretical toolbox. This is highly unfortunate because if we leave theory development to those who currently define themselves as theorists we will see little progress in the type of theory we are aiming for. In my vision of the future, analytical sociologists also routinely publish in journals like *Sociological Theory* and there they bring to the fore the general theoretical core of the matters they study in the form of theoretical models of social processes that bring about different types of macro outcomes.

Another consequence of being embedded in a somewhat foreign theoretical culture is that it forces us to engage in debates that often are entirely unproductive from our point of view. Two influential social theory traditions, that also are featured in this volume, critical realism and pragmatist theory, can be used to illustrate my point. Although there are some surface similarities between analytical sociology and these traditions, such as the concern with mechanisms, the analytical sociology agenda in other respects is so far removed from their agendas that the prospect for a productive dialogue is minimal.

In the case of critical realism, as emphasized by Di Iorio and León-Medina (this volume), most of these difficulties can be traced to their rather obscure ontological commitments. One aspect of this is their firm commitment to Bhaskar's (1978) view of reality as composed of three separate domains, the empirical, the actual, and the real. Since they assume that mechanisms operate at the empirically non-observable domain of the real, they exclude by definition the possibility of constructing the kind of empirically fine-tuned mechanism models that analytical sociology aims for. Another ontological assumption that hampers the prospect for a fruitful dialogue is their assumption that social structures have causal powers that are autonomous of the individuals who are embedded in them. As emphasized by Di Iorio and León-Medina, this assumption introduces a "mystery gap" between micro behaviors and macro outcomes that closes the door to any serious dialogue with analytical sociology about

Peter Hedström - 9781789906851

generative social processes. Macro outcomes often are more difficult to explain than isolated micro events, particularly when numerous interacting individuals are part of the process, but the fact that something is complex and difficult to explain does not mean that it is inexplicable or that emergent social structures are autonomous from their micro bases.

Pragmatist theory is not at all as marred by obscure ontological commitments as critical realism is, and the work of pragmatists such as Gross (2009) and Norton (this volume) who discuss the relationship between pragmatist theory and analytical sociology, is clear and eminently understandable. In his article on a pragmatist theory of social mechanism, Gross (2009: 375) argues as follows:

> Sociology should aim to identify the main social mechanisms by which cause and effect relationships in the social world that are of moral, political, or intellectual importance come about. This entails breaking complex social phenomena into their component parts to see how aggregations or chains of actors employing habits to resolve problem situations bring about systematic effects.

Although most analytical sociologists may not feature "habits" and "problem situations" as centrally as Gross does in a general programmatic statement, I assume that most would find his statement perfectly agreeable. That habits of the mind as well as habitual behavior can be important for explaining macro outcomes seems entirely uncontroversial, particularly when allowing for moments of "creativity" at which such habitual patterns are altered. It is more unclear what explanatory purchase there is in framing the analysis in this pragmatist cloak. Simply introducing the concepts and the perspective will not expand our theoretical toolbox or offer any new insights into complex social processes.

I am open to the possibility that pragmatist ideas may produce important additions to our theoretical toolbox, but neither Gross nor Norton make any serious move in that direction. Their apparent lack of interest in clearly and precisely detailing how different types of macro outcomes are brought about is rather typical for mainstream social theorists. As suggested by Kroneberg and Tutic (this volume), most social theorists are exclusively concerned with developing *conceptualizations* of actions, practices, etc., and in that respect they differ considerably from analytical sociology. The view of sociological theory as a toolbox of sometimes-true theories implies a much more pragmatic and flexible attitude toward the conceptualization of actions and practices. There is no general conceptual scheme that all analyses must make use of. Instead, a proposed mechanism is judged exclusively in terms of whether it is clearly and precisely articulated, whether it is empirically plausible, and whether it leads to new or more precise models of important social processes.

To what extent has analytical sociology benefited from dialogues with these traditions? Not much, I'm afraid. For a field to have a fertile impact on the analytical agenda, it must at least share our commitments to clarity and precision, to empirical realism and to generative explanations of social change. If none of these requirements are met, dialogues are likely to be distractive rather than constructive. For these reasons I am highly skeptical of the idea of developing an "analytical sociology in the pragmatist mode" as suggested by Norton (this volume).

What is needed in order to show that a pragmatist move has something to offer is not further conceptual elaborations but tangible evidence, preferably in the form of clearly articulated models of different types of social processes. In order for pragmatist statements to be taken seriously they must do something similar to what Goldberg and Stein (2018) did for our understanding of the diffusion of culture. That is, to include some of their core ideas in a formalized model of a social process, and tangibly show that this will enrich our

Peter Hedström - 9781789906851

understanding of the ways in which macro outcomes are brought about (see also Manzo 2021 this volume).

As I see it, analytical sociology gains much of its strength from being open in conceptual matters, from not being tied down by any general social theory, and for being wholeheartedly focused on explaining important macro outcomes in the sense of showing how the outcomes were brought about. I fear that a closer dialogue with traditional social theorists will simply lead to endless discussions about purely conceptual matters. The philosophy of science community is likely to offer much more productive dialogues as far as meta-theoretical matters are concerned, since they tend to share our commitment to clarity and precision, and interactions with empirically oriented sociologists—experimental, quantitative, and qualitative—are likely to be much more conducive for the development of new middle-range theories than dialogues with social theorists.

4. DO WE NEED INTENTIONAL EXPLANATIONS?

Individuals' actions are of central importance for analytical sociology since the social processes we are particularly concerned with are those that steer a group of individuals into acting in such a way that they collectively bring about the type of outcome to be explained.[3] The fact that individuals' actions are part of the process often has been understood as necessitating two different types of explanations (e.g. Elster 1983): First, we must intentionally explain why individuals do what they do and, second, we must causally explain why they bring about the macro outcomes they do when they act as they do.

In *Dissecting the Social* I also emphasized the importance of firmly anchoring all explanations of social outcomes in intentionally explained behavior. I argued that "intentional explanations are important for sociological theory because… they make the act 'understandable' in the Weberian sense of the term.… [I]n order for a theory [of the social] to be explanatory it must consider the reasons why individuals act as they do."

I am now considerably more skeptical to the feasibility and desirability of such an approach, and this is for three interrelated reasons:

1. In order to assess the correctness of a proposed intentional explanation, we must have access to information on the mental states of those whose behavior we wish to explain, but such information is scarcely ever available.
2. Our theories of action are not nearly powerful enough to serve as a substitute for the missing empirical information on the mental states of the individuals.
3. If we nevertheless were to use Elster's double-explanatory approach, we would commit what Elster (2009) himself has referred to as the fallacy of "excessive ambitions", and this is something we should avoid at all costs.

Let me expand a bit on each of these three points.

Following Davidson (1980), mental states such as beliefs and desires can be said to cause our actions in the sense of providing reasons for them, and knowing an individual's mental

[3] Please note that I use the terms *action* and *behavior* as synonyms, and as referring to the behavior of an individual, whether or not that behavior is guided by an intention.

Peter Hedström - 9781789906851

states allows us to understand why the individual did what they did. An intentional explanation of an individual's behavior refers to the specific constellation of mental states that motivated the behavior in question.

Intentional explanations are appealing because they provide subjective understanding. They are difficult to use appropriately, however, since we scarcely ever know the reason that motivated the behavior of another individual. I think it is fair to say that sociologists have not fully appreciated this problematic aspect of intentional explanations. Merton's (1936: 897) discussion of how to impute reasons to individuals is a case in point:

> Ultimately, the final test [of the imputation] is this: does the juxtaposition of the overt action, our general knowledge of the actor(s) and the specific situation and the inferred or avowed purpose 'make sense,' is there between these, as Weber puts it, a 'verständlicher Sinnzusammenhang?' If the analyst self-consciously subjects these elements to such probing, there is substantial probability that his conclusion in respect to purpose is not too far afield in the majority of instances. (p. 897)

This way of seeking to understand other individuals' behavior by placing us in their shoes is common practice in everyday life. The procedure also may be sufficiently reliable for some social-scientific work, particularly if we only need to impute reasons to a handful of persons whose life histories and outlooks on the world are very similar to our own. What Merton and likeminded analysts have failed to appreciate, however, is the considerable heterogeneity that exists in individuals' motivations. Some may do X because of belief B, while others may do X because of belief C or desire D, and the same individual may do X for entirely different reasons at different points of time.

Particularly in our kind of research where the focus is on the macro outcomes that *numerous* interacting individuals gradually bring about, this verstehen-based procedure is likely to be highly error prone. We can know how the individuals act, in the current as well as in the past, we can know the socio-demographic properties of the individuals themselves as well as the social settings in which they are embedded, and we can know their relationships to others and how the others around them acted in the past, but normally we will not have access to what goes on in their minds and how their mental states change as the social process unfolds. For this reason, we cannot know the specific constellation of mental states that made the individuals' act as they did, and this would seem to preclude the use of intentional explanations in our kind of research. Although there exist ways of eliciting individuals' reasons and emotions in lab settings, or via survey instruments, such procedures are not easy to adapt to large-scale observational studies. This is particularly so if the proposed explanation hinges on specific changes in the individuals' mental states during the process being analyzed.[4]

Schematically the situation may be described as follows, where **A** is the action of an individual, **M** is the individual's relevant mental states at the time of acting, and **S** are various observable social characteristics of the individual and the individual's social environment likely to affect the individual's mental states and actions:

$$\mathbf{S} \rightarrow \mathbf{M} \rightarrow \mathbf{A}$$

[4] Since we cannot elicit individuals' mental states on the fly, the only option would seem to be to use time-stamped digital data, when available, as an indicator on the individuals' reasons and emotions at the time when they acted. However, such data rarely is available, and often does not map properly onto our theoretical constructs.

Peter Hedström - 9781789906851

An intentional explanation is focused on the **M→A** part of the scheme. As with any other type of explanation, an intentional explanation can be correct or incorrect. The explanation is correct if it refers to the mental states that actually motivated the individual to behave as she/he did, and it is incorrect if it refers to the wrong set of mental states. Since we do not know **M**, and since many different constellations of **M** can motivate the same **A**, the prospect for developing correct intentional explanations seems meager indeed. The inaccessibility of **M** also casts doubt onto sociological research programs that hinge on details about **M** and the **M→A** link, such as Esser's frame-selection approach (e.g. Esser 2009).

If we had a sufficiently strong theory of action, we could partly circumvent this problem by using this theory to impute the missing **M→A** component. This was the position taken by Popper (1961, 1994) when proposing his form of *situational analysis*. Inspired by economic theorists and their methodology of constructing explanations of economic behavior by imputing a motive to the individuals, and then derive what their optimal choices would be in the circumstances in which they found themselves, Popper argued that a similar approach would also be adequate for most kinds of social explanations (Hoover 2017; Hedström et al. 1998). Popper seems to have believed that this assumption was theoretically innocent and an almost empty principle that he nevertheless found useful because it allowed us to "pack or cram our whole theoretical effort, our whole explanatory theory, into an analysis of the situation" (1961: 359).

Similar considerations are likely to explain why so many social scientists were attracted by rational choice theory. Although we cannot, but in rare circumstances, know what motivates individuals to do what they do, rational choice theory provided a coherent narrative that could be used as a substitute for this missing information. This use of rational-choice theory was perhaps most clearly seen in the work of Gary Becker and similarly minded theorists who sought to develop a new foundation for the social sciences, not by painstaking empirical research into the true motives of individuals' behavior, but by proposing general and model-based narratives that provided rational-choice–based interpretations of various kinds of behavior (e.g. Becker 1976, 1981; Coleman 1990). What Becker and his followers showed is that it is possible to come up with coherent rational narratives that fit even the most puzzling and seemingly irrational kinds of behavior, particularly if the narratives are free to ignore known empirical facts about the behavior of individuals.[5] Inadvertently, Becker et al. thereby also demonstrated that being able to tell a coherent narrative about a possible causal process is far from sufficient for an acceptable explanation to be at hand. In order for a narrative to be explanatory, it has to be firmly anchored in known facts about the acting individuals and their social settings.[6]

---

[5] During my years at the University of Chicago, the famous Becker–Coleman seminar was at its height, and this seminar in many respects was an ideal type of this kind of rational-narrative approach. The typical seminar format was that an invited speaker presented some research and proposed an explanation, not seldom of a non-rational-choice kind. Very quickly the regular seminar participants interrupted the speaker and proposed numerous, often very ingenious alternative rational choice explanations of the findings. The seminars were intellectually extremely stimulating, but obviously they had more in common with purely intellectual games of Herman Hesse's type of glass-bead game than with an empirically grounded social science proper.

[6] Another serious problem with these kinds of post hoc accounts based on little or no empirical data is their lack of predictive power. While, as has been discussed, there is more to a good explanation than being able to predict if a proposed explanation fails miserably in predicting the kind of phenomena it is supposed to explain, it is a sure sign that something is seriously wrong with the proposed explanation (see Watts 2014 for an interesting discussion of this point).

Peter Hedström - 9781789906851

Although these kinds of just-so stories where the theorist makes up the acting individuals's motives may provide a sense of understanding, this understanding is illusionary as long as the story is factually incorrect. If we nevertheless push ahead and act as if the micro foundations were correct, we commit what Elster (2009) has referred to as the fallacy of "excessive ambitions". He introduced this notion in a rather damaging critique of Bates et al.'s (1998) attempt to explain complex historical developments in rational-choice terms (Elster 2000), and he argued correctly that this sort of explanatory hubris not only leads to poor and misleading explanations but that an acceptance of such principles is likely to hinder the development of serious social science research.

Rather than seeking excessively precise fictions, we should aim to develop theories applicable to the real-world settings that we are analyzing. Such a modest and realist strategy characterizes some of the best theoretical work in the discipline. Take Merton (1968) and his theory of the Mathew effect as an example. This theory specifies the details of a cumulative advantage process in which success breeds further success, in academia as well as elsewhere (see also Lynn and Espy, this volume). Merton never specified any precise model of the decision calculus, nor did he propose any elaborate **M**-story. Doing so would be possible, of course, but if the resulting model did not properly describe what transpired in the case at hand, the formalization would render the account non-explanatory. The situation is similar with Schelling's (1971) classic segregation model that has been an important source of inspiration for analytical sociology. What matters for the aggregate segregation patterns that emerge from the process that Schelling analyzes is *how* individuals react to the behavior and properties of their neighbors, not *why* (in the intentional or mentalistic sense) they react as they do. That is, the segregation dynamics will be the same irrespective of the reasons the individuals do what they do. The crucial aspect that gives the process its unique properties is not related to the **M** or the **M→A** component, but is exclusively driven by the interaction patterns as such, for instance, the extent to which the **A** of one individual influences the **A** of another individual.[7] In both of these cases, more details clearly could be specified and might also be desirable, but not excessive precision that simply amounts to precisely stating and assuming to be true what is untrue, and which thereby would deprive the theory of its explanatory power.

What this discussion suggests is that it is important to make a clear distinction between explanations of actions, on the one hand, and intentional explanations (or action theories founded on such explanations), on the other. In my view, the **M→A** information problem is of such significance that we should avoid intentional explanations in all but the utterly rare cases where we have access to dependable information on **M**. Since the essential **M**-information usually is inaccessible to us, we can neither test the assumed details of the **M→A** link nor

[7] One clarification is called for. If the actors in the Schelling model move when 20 per cent of the neighbors are of the other kind rather than 60 per cent, this will affect the segregation patterns that emerge. One interpretation of this behavioral difference could be that the former group has stronger segregationist preferences than the other, but as long as we do not have any independent information on their **M**-states, this is only a guess. This behavioral difference between the groups could, equally well, arise if there were no differences in their segregationist preferences. It could, for example, arise if the 60 per cent group owned its housing and, therefore, was concerned with how the inflow of the other kind of actors may come to affect their property values, while the other group lived in rental properties and, therefore, was entirely unconcerned about property values. All we know for sure is that we have two groups that differ in their thresholds for leaving their neighborhoods, and that this difference in how they react to others affects the aggregate dynamics. We do not know anything about **M** and the **M→A** link, and my point is that we should avoid making our explanations hinge on causal narratives that we do not have firm empirical support for.

Peter Hedström - 9781789906851

fine-tune our theoretical accounts on the basis of detailed empirical analyses of the **M→A** link. As a result, proposals about the **M→A** link will be unreliable and explanations founded upon them are likely to be highly error prone.

As the examples of Merton and Schelling illustrate, the fact that intentional explanations are rarely possible to use is not necessarily of much importance for us. The reason for this is that most of the social processes that we are concerned with are not dependent on motivational details, but are driven by **A**-based interaction patterns as such. Thus, in many situations, what is essential for explaining macro outcomes is not to have a model of the individuals' deliberations, but to have an accurate model of social interactions and their effects.

The lack of information on **M**, combined with the premium attached to intentional explanations within the discipline, has led to a proliferation of just-so stories with little or no empirical support. This practice comes into conflict with what I consider one of the main pillars of analytical sociology; that our explanations should always refer to processes actually responsible for the outcomes that we seek to explain. If we do not have information on **M** it seems obvious to me that we should not pretend to know what kind of **M→A** link is at work, but instead frame the explanation exclusively in **A** and **S** terms. But equally obvious, of course, if we happen to have access to relevant **M** information for the individuals involved in the process, we should make full use of it in our explanation.[8]

5. THE STRUCTURE OF SOCIAL INTERACTION

For all the reasons mentioned, unless we have access to reliable **M** information, we are well advised to focus our theoretical and explanatory efforts on the analysis of the social situation and leave theorizing about individuals' minds to those who have specialized knowledge in such matters; i.e. psychologists and cognitive scientists (see also Hedström and Ylikoski 2014). That is to say, we should cater to our own strengths and focus, theoretically as well as empirically, on how the social situations in which individuals are embedded affect their behavior. Since other individuals are at the core of these social situations, this implies detailed analyses of social interactions, the ways in which they are structured, and how they affect individuals' behavior and the macro outcomes the individuals collectively bring about.

Michael Mäs's chapter on interactions (this volume) calls attention to Lopez-Pintado and Watts' (2008) notion of *influence-response functions*, and he argues that such functions can

[8] The position taken here differs in important ways from that taken in my previous writings on the so-called DBO model (Hedström 2005). In our everyday life we make constant use of the DBO scheme or similar folk-psychological ideas to make sense of the behavior of others. The scheme is flexible and well adapted for such everyday needs, but it does not work as well as a theory of action. This is most clearly seen in its failure to predict the future behavior of individuals. As mentioned, although there is more to a good explanation than predictive success, if a theory of action fails miserably in predicting the behavior it is supposed to explain, it is a sure sign that something is seriously wrong with the theory (Watts 2014). It is this tension between the psychological appeal of intentional emphatic explanations, on the one hand, and the realist dictum that proper explanations must refer to the correct causal structure, on the other, that has led me to the current position. In Hedström (2005) such considerations led me to express great disbelief in rational choice theories, and in Hedström and Ylikoski (2014), on similar grounds, we questioned the very idea of basing explanations on sociological theories of action. In this chapter, I take the final step and argue for going fully naturalistic; i.e. to give up on the privileged role of intentional explanations, and instead adopt an explanatory account of individual behavior that is directly influenced by our empirical research.

Peter Hedström - 9781789906851

serve to coalesce many seemingly disparate ideas related to the structure of social interaction. Influence-response functions may indeed provide a unifying lens for analytical sociology because they refer to social interactions as such, and they show how an individual's behavior is affected by the behavior of others and, thereby indirectly, how the structure of social interaction is likely to affect the macro outcomes the individuals bring about. The core idea behind such functions has been around for a long time, but that does not diminish its importance.[9]

So, what is an influence-response function precisely? Let me explain with reference to a study of the spatial diffusion of Swedish social movements at the turn of the nineteenth century (Hedström 1994). In this study I took my departure in Granovetter's (1978) threshold model and used influence-response ideas to extend it so that it also could consider the networks in which the individuals were embedded. Central to the influence-response approach is that we have a *n* by *n* matrix **S**, which describes the structure of social interaction between *n* individuals. In my spatial diffusion study, the entries of **S** were inverted spatial distances between the actors, but more generally **S** refers to the social network the individuals are embedded in, or to the extent to which individuals pay attention to other individuals' behavior. In addition to this social-interaction matrix, we have a *n* by *1* column vector, **B**, which identifies those who had adopted the behavior in question at the previous point in time; the influence-response function is given by the *n* by *1* column vector **I**:[10]

$$\mathbf{I} = \mathbf{S} \times \mathbf{B}$$

The entries of **I** constitute the exposure part of the influence-response function and by relating this exposure to each individual's behavior at the next point in time we have a function that shows how individuals respond to the influences emanating from others. As emphasized, individual heterogeneities are common and important for the collective outcomes the acting individuals are likely to generate. Some individuals, for example, are more susceptible to social influence than others. Within an influence-response framework, such heterogeneities can be handled in terms of thresholds in Granovetter's (1978) sense of the term and, as shown in Hedström (1994), individually specific thresholds can be empirically estimated with standard statistical techniques.

I used Schelling's segregation model to bring home the point that the macro outcomes that interacting individuals bring about often are exclusively driven by *how* they react to each other,

---

[9] As far as I know, the core idea was introduced into the social sciences by the social physicist Stewart (1941, 1947) in the 1940s (see also Duncan 1957), and similar ideas are central to Latané's (1981) social-impact theory (see also Nowak et al. 1990), Centola's (2018) work of the spread of behavior, Friedkin's work on network-based social influence (e.g. Friedkin and Johnsen 1999), Land et al.'s (1991) analysis of church attendance, and my own work on contagious collectivities (Hedström 1994; Hedström et al. 2000), just to mention a few.

[10] Assume, for example, that we study a minimal social system comprised of only three individuals and that

$$\text{their } \mathbf{S} \text{ matrix is equal to } \begin{bmatrix} 0 & .9 & .8 \\ .1 & 0 & .2 \\ .4 & .4 & 0 \end{bmatrix}$$

These entries suggest that the first individual, in row 1, pays a lot of attention to the other two individuals, that the second individual, in row 2, pays little attention to the others, and that the third individual, in row 3, occupies a position in between the other two. If the first two individuals, but not the third individual, exhibited the behavior in question at the previous time period, their behavior is given by this

$$\mathbf{B} \text{ vector } \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \text{ and the } \mathbf{I} \text{ vector then is equal to } \begin{bmatrix} .9 \\ .1 \\ .8 \end{bmatrix}$$

Peter Hedström - 9781789906851

and not by *why* (in the intentional sense) they react as they do. What Pintado and Watts and Mäs show is that this conclusion can be generalized to wide classes of behavior.

Pintado and Watts focused on three types of influence-response functions: (1) monotonically increasing functions where the probability of an individual adopting a behavior increases monotonically with the number of others who already have adopted it; (2) monotonically decreasing functions where the probability of an individual adopting a behavior decreases monotonically with the number of others who already have adopted it; and (3) non-monotonic functions where the probability of adoption increases (decreases) up until a certain number of others have adopted the behavior after which it decreases (increases). They show that substantively very different kinds of behaviors can have very similar influence-response functions and, to the extent that this is the case, they will give rise to similar macroscopic patterns.

Instead of investing time and effort into developing stories about an unobserved **M→A** component, how it may vary between different individuals, and how it may change during the course of a social process, there are strong reasons to focus on **S** and **A** instead, and adopt and further develop the influence-response approach. This approach steers our focus to those matters on which we can speak clearly, and which clearly matters for the macro outcomes we seek to explain. All key aspects of the influence-response functions can be studied empirically, and the form of the functions can be estimated with the kind of empirical data that contemporary sociologists have access to (see Hedström et al. 2000; Keuschnigg et al. 2018).[11]

6. WHAT'S NEXT? SOME CONCLUDING REMARKS

As a way of conclusion, let me concisely summarize the main message of the chapter in terms of four particularly important priorities for the future:

1. Invest more time and effort into developing the toolbox of middle-range theories.
2. Avoid intentional explanations of behavior, unless reliable data on the individuals' mental states is available.
3. Pay less attention to mainstream social theory and more attention to fields that share our commitments to clarity, empirical realism, and generative explanations of social change.
4. Embrace the computational social science revolution.

As noted, the development of our theoretical toolbox has not kept pace with the development of our applied work. In the spirit of Thomas Schelling's message at the 1996 social mechanism conference in Stockholm, we need to systematize sociological theory by creating better order in our theoretical toolbox, and by identifying important areas where the existing toolbox is shallow or empty. Many parts of the theoretical toolbox would benefit from close attention, but in the light of the discussion it appears particularly important to systematize and deepen our understanding of influence-response functions. Since relevant work on such functions is scattered into many different areas, one important task would be to

[11] But once again, I am not arguing in favor of a behavioristic approach that denies the existence of mental states or their possible causal efficacy. What I am arguing against is just-so storytelling about the **M→A** component. If we have access to **M** information, we should make use of it and analyze the role that different aspects of the **M→A** component played in bringing about the macro outcome we seek to explain.

Peter Hedström - 9781789906851

coalesce this literature, show its common denominators, and identify what we know about the mapping between types of social situations and types of influence-response functions. In addition, in order for influence-response functions to take a central role in analytical sociology, extensive work is needed in order for us to fully understand the types of macro patterns that different types of influence-response functions give rise to. Using the terminology introduced, we need to know how the macro patterns that emerge depend on how the entries of **S** are structured, the initial entries of **B**, and the form of the function relating **I** to **B**. Similarly, we must understand what role individual heterogeneity may play in these processes. Such heterogeneities can arise in the **S** matrix in terms of clustering, but also in the link between **I** and **B**, since some individuals may be become more likely to adopt the behavior in question when many others have done so, while others may shy away from behaviors that have become "too common", and it is unclear how different mixtures of such groups may affect the collective outcomes they jointly bring about.

Let me end by saying a few words about the emerging field of computational social science (CSS) and its importance for analytical sociology. The reason that I have not discussed CSS is not that these developments are of marginal importance, but that we have already discussed their importance elsewhere (see Keuschnigg et al. 2018; Jarvis et al. (forthcoming)). As emphasized there, CSS has the potential to accomplish for sociology what the introduction of econometrics did for economics in the past half century, i.e. to provide relevant data and data-analytical tools needed for rigorously addressing the core questions of the discipline. As pointed out in several of the chapters of this volume, the kind of survey data that has dominated the discipline for several decades, and the statistical tools used for analyzing such data, are poorly equipped to handle analyses of networked social systems and their dynamics. During the last decade, computationally intensive statistical methods appropriate for handling data on large-scale populations of interacting individuals have emerged at a fast rate, as well as new and improved computational methods for analyzing large corpuses of text. These kinds of developments are important since they enable us to reduce the gap between empirical and theoretical work, and thereby allow empirical research to have more bearing on sociological theory than what has been possible in the past.

These developments are important not only for observational but also for experimental studies. Salganik et al.'s (2006) web-based experiments on artificial cultural markets represent a new type of macro-sociological experiments (Hedström 2006). Their experimental design breaks with older experimental traditions by randomizing participants into separate social systems that each serve as a unit of analysis. The focus is on the macro outcomes that the individuals in the different social systems collectively bring about, and for this reason this type of macro experiment is likely to be of considerable importance for the future of analytical sociology.

REFERENCES

Abbott, A. (1998), "The casual devolution", *Sociological Methods and Research*, **27**: 148–81.

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal and Barry R. Weingast (1998), *Analytical Narratives*, Princeton, NJ: Princeton University Press.

Bearman, P.S., J. Moody and K. Stovel (2004), "Chains of affection: The structure of adolescent romantic and sexual networks", *American Journal of Sociology*, **110**: 44–91.

Becker, G.S. (1976), *The Economic Approach to Human Behavior*, Chicago, IL: University of Chicago Press.

Peter Hedström - 9781789906851

Becker, G.S. (1981), *A Treatise on the Family*, Cambridge, MA: Harvard University Press.
Bhaskar, R. (1978), *A Realist Theory of Science*, Hassocks, Sussex: Harvester Press.
Bruch, E.E. and R.D. Mare (2006), "Neighborhood choice and neighborhood change", *American Journal of Sociology*, **112**: 667–709.
Centola, D. (2018), *How Behavior Spreads*, Princeton, NJ: Princeton University Press.
Coleman, J.S. (1964), *Introduction to Mathematical Sociology*, New York: Free Press of Glencoe.
Coleman, J.S. (1990), *Foundations of Social Theory*, Cambridge, MA: Harvard University Press.
Davidson, D. (1980), *Essays on Actions and Events*, Oxford, UK: Clarendon Press.
Duncan, O.D. (1957), "The measurement of population distribution", *Population Studies*, **11** (1): 27–45.
Elster, J. (1983), *Explaining Technical Change: A Case Study in the Philosophy of Science*, Cambridge, UK: Cambridge University Press.
Elster, J. (2000), "Rational choice history: A case of excessive ambition", *American Political Science Review*, **94** (3): 685–95.
Elster, J. (2009), "Excessive ambitions", *Capitalism and Society*, **4** (2).
Esser, H. (2009), "Rationality and Commitment: The Model of Frame Selection and the Explanation of Normative Action", pp. 207–30, in M. Cherkaoui and P. Hamilton (eds), *Raymond Boudon: A Life in Sociology*, Oxford: The Bardwell Press.
Friedkin, N.E. and E.C. Johnsen (1999), "Social influence networks and opinion change", *Advances in Group Processes*, **16**: 1–29.
Goldberg, A. and S.K. Stein (2018), "Beyond social contagion: Associative diffusion and the emergence of cultural variation", *American Sociological Review*, **83** (5): 897–932.
Goldthorpe, J.H. (2021), *Pioneers of Sociological Science: Statistical Foundations and the Theory of Action*, Cambridge: Cambridge University Press.
Granovetter, M.S. (1978), "Threshold models of collective behavior", *American Journal of Sociology*, **83**: 1420–43.
Gross, N. (2009), "A pragmatist theory of social mechanisms", *American Sociological Review*, **74** (3): 358–79.
Hedström, P. (1994), "Contagious collectivities: On the spatial diffusion of Swedish trade-unions, 1890–1940", *American Journal of Sociology*, **99**: 1157–79.
Hedström, P. (2005), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge, UK: Cambridge University Press.
Hedström, P. (2006), "Experimental macro sociology: Predicting the next best seller", *Science*, **311** (5762): 786–7.
Hedström, P. and P. Bearman (eds) (2009), *The Oxford Handbook of Analytical Sociology*, Oxford, UK: Oxford University Press.
Hedström, P. and G. Manzo (2015), "Recent trends in agent-based computational research: A brief introduction", *Sociological Methods and Research*, **44** (2): 179–85.
Hedström, P., R. Sandell and C. Stern (2000), "Mesolevel networks and the diffusion of social movements: The case of the Swedish Social Democratic Party", *American Journal of Sociology*, **106** (1): 145–72.
Hedström, P. and R. Swedberg (eds) (1998), *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge, UK: Cambridge University Press.
Hedström, P., R. Swedberg and L. Udehn (1998), "Popper's situational analysis and contemporary sociology", *Philosophy of the Social Sciences*, **28** (3): 339–64.
Hedström, P. and P. Ylikoski (2010), "Causal mechanisms in the social sciences", *Annual Review of Sociology*, **36**: 49–67.
Hedström, P. and P. Ylikoski (2014), "Analytical Sociology and Rational Choice Theory", in G. Manzo (ed.), *Analytical Sociology: Actions and Networks*, Chichester, UK: Wiley.
Hoover, K.D. (2017), "Situational Analysis", pp. 182–90, in L. McIntyre and A. Rosenberg (eds), *The Routledge Companion to Philosophy of Social Science*, London: Routledge.
Jarvis, B., M. Keuschnigg and P. Hedström (forthcoming), "Analytical sociology amidst a computational social science revolution", in U. Engel et al. (eds), *Handbook of Computational Social Science*, Routledge.
Keuschnigg, M., N. Lovsjö, and P. Hedström, (2018), "Analytical sociology and computational social science", *Journal of Computational Social Science*, **1** (1): 3–14.
Land, K.C., G. Deane and J.R. Blau (1991), "Religious pluralism and church membership: A spatial diffusion model", *American Sociological Review*, **56** (2): 237–49.
Latané, B. (1981), "The psychology of social impact", *American Psychologist*, **36** (4): 343–56.
Lopez-Pintado, D. and D.J. Watts (2008), "Social influence, binary decisions and collective dynamics", *Rationality & Society*, **20** (4): 399–443.
Manzo, G. (2007), "Comment on Andrew Abbott", *Sociologica*, **2**. doi: 10.2383/24752.
Manzo, G. (2014), "Data, generative models, and mechanisms: More on the principles of analytical sociology", pp. 1–52, in G. Manzo (ed.), *Analytical Sociology: Actions and Networks*, Chichester, UK: Wiley.
Manzo, G. and A. van de Rijt (2020), "Halting SARS-CoV-2 by targeting high-contact individuals", *Journal of Artificial Societies and Social Simulation*, **23** (4).
Merton, R.K. (1936), "The unanticipated consequences of purposive social action", *American Sociological Review*, **1**: 894–904.

Peter Hedström - 9781789906851

Merton, R.K. (1948), "The Self-Fulfilling Prophecy", pp. 475–90, in Robert K. Merton (ed.), *Social Theory and Social Structure*, New York: The Free Press.
Merton, R.K. (1967), "On sociological theories of the middle range", pp. 39–72, in Robert K. Merton (ed.), *On Theoretical Sociology*, New York: The Free Press.
Merton, R.K. (1968), "The Matthew effect in science: The reward and communication systems of science are considered", *Science*, **159** (3810): 56–63.
Nowak, A., J. Szamrej and B. Latané (1990), "From private attitude to public opinion: A dynamic theory of social impact", *Psychological Review*, **97** (3): 362–76.
Popper, K.R. (1961), *The Poverty of Historicism*, London: Routledge.
Popper, K.R. (1994), "Models, instruments, and truth: The status of the rationality principle in the Social Sciences", pp. 154–84, in K.R. Popper (ed.), *The Myth of the Framework: In Defence of Science and Rationality*, London: Routledge.
Salganik, M.J., P.S. Dodds and D.J. Watts (2006), "Experimental study of inequality and unpredictability in an artificial cultural market", *Science*, **311** (5762): 854–6.
Schelling, T.C. (1971), "Dynamic models of segregation", *Journal of Mathematical Sociology*, **1**: 143–86.
Stewart, J.Q. (1941), "An inverse distance variation for certain social influences", *Science*, **93** (2404): 89–90.
Stewart, J.Q. (1947), "Empirical mathematical rules concerning the distribution and equilibrium of population", *Geographical Review*, **37** (3): 461–85.
Watts, D.J. (2014), "Common sense and sociological explanations", *American Journal of Sociology*, **120** (2): 313–51.

Peter Hedström - 9781789906851

Index

Note: Page numbers followed by n and number represent footnotes and note number.

abstract simulation models 187
The Achieving Society (McClelland) 50
action-based models (ABMs) 187
action-formation mechanisms 86
action theory 32–5, 175–8, 186–8
  fundamental challenge of 188–9
activity 435
A-facts, Coleman diagram 50–1
agent-based computational modeling (ABCM) 25–31, 71, 314, 453–5
  antecedents of 91
  bird's eye view on 455–6
  model analysis 459–64
  realistic model of relocation decision making 464–9
  social context 456–7
  types of 187
  on unintended residential segregation 456–69
agents
  as ABCM's element 455
  heterogeneity of 58
  pre-formed 166
Albert, Hans 82
algorithmic extremism 354, 354n1
Allais Paradox 194
alliances
  political 18–19, 25
alternative influence-response function 211–12
American pragmatism 69
analytical action theory, *see* action theory
analytical Marxism 64, 73–5
  analytic philosophy and 76–7
Analytical Marxism: Studies in Marxism and Social Theory (Roemer) 73
analytical sociology (AS) 64
  analytical style of 338–40
  analytic philosophy relationships with 69–71
  antecedents of 80–1, *see also* structural individualism
  characteristics of 491
  complexity perspective on 109–16
  core of 490–3
  critical effervescence (2007–2014) 4
  and cultural sociology, *see* cultural sociology
  individuals' actions and 496–500
  intentionalism of 156–7
  manifestos 2, 4, 7–8, 12, 22
  methods, digital data and 357–61
  models, *see* models
  non-foundationalism of 138–9
  principles 5–6
  programmatic variations in substantive works 7–10
  programs and sub-programs in 2–10
  re-formulation and new reactions (2014–) 5–6
  Social Stratification Research, *see* Social Stratification Research (SSR)
  theoretical developments 493–6
analytic philosophy 64, 65–73
  and analytical Marxism 76–7
  assessment 75–7
  dividends to analytical sociology 72–3
  relationships with analytical sociology 69–71
anglophone philosophy 69, 74
antecedents
  of agent-based modeling 91
  of analytical sociology 80–1, *see also* structural individualism
anti-scientific 137
a priori 17
archival data 337–8
  digital archives 345–8
  sources 341–8
  tangible archives 341–5
arguments
  causal 264
  retroductive 148–50
  transcendental 147, 149
arrows, Coleman diagram 51–3, 59
asociological atomism 54–7
assimilative social-influence model 210–11
associative diffusion 21–2, 28, 164
assortative mating 487
assurance dilemma (AD) 419, 420
asymmetric cooperation norms 233, 239–41
attraction 435
attrition 404
audience valuation 295–7
average causal effect (ACE) 367, 368
average treatment effect (ATE) 367, 397

Gianluca Manzo - 9781789906851

Average Treatment on the Treated (ATT) 397
averaging-assumption 210
Axelrod's seminal model of the dissemination of culture 209
axiomatic decision theory 190

backfire effects 331–2
basic reproduction number 480
battle of the sexes 422
Becker, Gary 160
behaviour evaluation 164
belief-based interactions 315
betweenness centrality 326
between-subject design 407
B-facts, Coleman diagram 51
blockmodeling 222
Borges, J.L. 158
Borgian Hypercube 475, 483, 485
Boudon, Raymond 164, 249
bounded rationality 195–6
Breen-Goldthorpe model 122, 130, 131–2
Brenkert, George 73
Brenner, Robert 74
A Brief History of Analytic Philosophy (Schwartz) 66
"broken window" theory 160

Capital in the Twenty-First Century (Piketty) 486
cardinal utility theory 191–2
career-focused interactions 259
career-type datasets 341
Carnap, Rudolph 68
categorization error 479
causal arguments 264
causal criterion of reality 145
causal independence 277–8
causal interdependence 60n8
causality 178–9
  approaches to 394–411
  defined 394
  social mechanisms and 144–7
causal mediation effect 383–4
causal validity 400–1
cause and effect 228–30
centrality
  betweenness 326
  complex 326–8
  degree 326
  eigenvector 326
central nodes 256–7
Chain of Affection (Bearman, Moody and Stovel) 155
chance 298–302
  contingency of 272–3
  defined 273
chaos game 105–7
checkerboard model 59
chicken dilemma (CD) 423
choice homophily 309–10
Civil Rights Movement 325
claims
  validation of 277–80
cluster, form of RCT 394
clustering 223
coalition norms, group pressure as 244–7
Coleman, James 49, 112, 249
Coleman's boat/bathtub 102, 110, 166, 476–7
  arrows 51–3, 59
  feedback loop 59–60, 59n6
  and heterogeneity of agents 58
  interesting reasons 49
  macro-to-micro influences 54–7
  micro-macro transition 58–61, 86
  motivation to use 53–4
  nodes 50–1
  overview 49–50
  structural individualism 86–7
  theory of action 57–8
collective intelligence 475
collective intentionality 146
collective properties, of structured entities 145
collider variable 373–5
complex contagion 322
  backfire effects 331–2
  countervailing influences 324–5
  diversity, power of 330–1
  influencers 325–6
  network bias 329–30
  process 311
  research on 323
  similarity, power of 330–1
  simple contagion *vs.* 322
  social tipping points 328–9
  surprising example 323–4
complexity, models 478–9
complexity research 100–1
  micro-macro problem in 105–8
complex network concepts 432
complex path length 326–8
complex system 105
  chaos game 105–7
  examples in various disciplines 108
  spontaneous stratification 107–8
computational model 37–8
  agent-based 25–31, 71
conditional uniform graphs (CUGs) 438–40
conditioning 372–3
confirmation, context of 68
conflict 422–4

Gianluca Manzo - 9781789906851

conjunctural causation 272
conjuncture 266
consolidation 308
  correlation as 487
"constellation of interests" 256
constraints, system-imposed 293–5
constraint satisfaction 164
construct validity 401
contagions 321–2
  complex, *see* complex contagions
  simple 322
contexts 249–50
  coupled, and social rationality 253–5
contingency 264–6
  of chance 272–3
  classification grid 268–71
  conditions and impacts 280–2
  empirical meaning of 265
  methodological *vs.* substantive 269–71
  nominal *vs.* real 268–9
  ontological claims 271–2
  open-ended conjunctures 274–7
  symptomatic 271
  validation prospects 277–80
contingent second wave property 482
contracted reductionism 249
Controlled Direct Effect (CDE) 381
cooperation dilemma 415–18
cooperation norms
  asymmetric 233, 239–41
  cultural norms and 241–4
  symmetrical 236–9
coordination dilemmas 418–22
coordination game 208, 209
coordination role, of norms 234–5
counterfactual causal model 367, *see also* potential outcome model
countervailing influences 324–5
coupled contexts, and social rationality 253–5
Cournot, Augustin 272–3
Cournot contingency 265
COVID-19 pandemic 315, 329, 330, 480–2
credibility 321
critical effervescence (2007–2014) 4
critical realism 135, 494
  experiments 148–51
  as ontology of deep structures of reality 135–8
  retrodiction 150–1
  retroduction 148–50
  simulation 148–51
cross-norm inhibition effect 21
crossover, form of RCT 394
cultural differentiation 29
cultural dope 156
cultural norms 241–4
cultural sociology
  analytical sociology *vs.* 158–64
  deviance as situational choice 158, 160–1
  overview 155–8
  veil as signal and self-binding 161–2
cultural transmission/differentiation 163–4
cumulative advantage (CA) 286–7
  audience valuation 295–7
  decoupling 291–302
  divergence 291–302
  endowments 291–3
  patterns of divergence 288–91
  research on 302–3
  self-selection 297–8
  system-imposed constraints 293–5

Dark Net 418
data
  archival, *see* archival data
  digital, *see* digital data
  network 437
data structures 340–1
decoupling 289–91
  explaining 291–302
degree centrality 326
descriptive norms 204
Desires-Beliefs-Opportunities (DBO) model 157, 175–8, 249–50, 254, 315–16, 500n8
determinism *vs.* intentionality 155–6
Dewey, John 69
digital archives 345–8
digital data 352–5
  and analytical sociology methods 357–61
  descriptive value of 356
  digital experimental data 354
  emojis 354
  opportunities and challenges 355–7
  relational data 353–4
  sensor data 353
  spatial data 353
  temporal data 353
  textual data 353
  variety of 352–3
  video data 353
digital experimental data 354
digital methods 352–5, 354
DiMaggio, Paul 156
directed acyclic graphs (DAGs) 364, 369–71
Direct Effect (DE) 381
discovery, context of 68
discrete shifts in divergence 287–9
discrimination 181
Dissecting the Social: On the Principles of Analytical Sociology (Hedström) 65, 69–70, 80, 226, 364, 490, 496

Gianluca Manzo - 9781789906851

divergence
  discrete shifts in 287–9
  explaining 291–302
  incremental 287–9
  patterns of 288–91
diversity
  models 475–7, 483–5
  power of 330–1
Diversity Prediction theorem 479, 480
downward causation 55, 146
Dual-Process Perspective (DPP) 196–200
Durkheim, Emile 100–3, 108–9, 204, 233
DyNAM model 447–8

East India Company (EIC) 339, 340
ecological validity 403–4
Education, Opportunity, and Social Inequality (Boudon) 81–2, 91
educational inequalities 122
effective reproduction number 481
eigenvector centrality 326
Ellsberg Paradox 194
Elster, Jon 64, 72, 73, 74, 76
emergent, system-level phenomenon 105
emojis 354
emotional excitement 322
empirical research, six mechanism-oriented pieces of 7
endowments 291–3
epidemiological models 163
epistemic fallacy 137
Epstein, Joshua 71
Erdös, Paul 222
Esser, Hartmut 82, 197
ethics
  normative 68
  philosophy of 67
European Network for Philosophy of Social Science 72
experimental effects 404–5
experimental realism 401–2
experiments 392–411
  design, types of 406–7
  field 392
  lab 392
  natural 392–3
  online 393, 407–11
  quasi-experiments 396
  random assignment 396–9
  types of 392–3
  validity, *see* validity
explananda, Merton award 13–18
explanans, Merton award 14–15, 18–22, 34
explanatory sociology 84n6
Exponential Random Graph Models (ERGMs) 441–3
externalities 206, 315
external validity 402–3

Facebook 324
factorial, form of RCT 394
fallibilism 138
feedback loop 59–60, 59n6
field experiments 392
formalizing sociological explanations 131–2
formal norms 234
Foundations of Social Theory (Coleman) 49, 250
founding stage (1996–2009) 2–3
French privateering networks 341–5
French Revolution 341
friendship networks
  ethnic segregation in 22
front door criterion 375–6

Gambetta, Diego 64
game-theoretic models
  conflict 422–4
  cooperation dilemma 415–18
  coordination dilemmas 418–22
  overview 414–15
  repetition 424–7
game theory 237
gas station theory 252
general theories 114–16
generative methodology 3
generativist theory of social explanation 65
Gershuny, Jonathan 166
Goldthorpe, John 120
Goodman, Nelson 69, 220
Google 331–2
Granger causation 395
Granovetter, Mark 223, 321, 328
Granovetter's seminal threshold model 208
group-level natural selection 129n5
group norms 233, 244–7
growth rates 287
guidelines, for analytical sociology's research program 40–5

The Handbook of Analytical Sociology (Hedström and Bearman) 226
health buddies 225
Hedström, Peter 65, 364
heuristics
  for micro-macro modeling 89–91
  negative 5
  positive 5

Gianluca Manzo - 9781789906851

Heuristics and Biases Program (Tversky and Kahneman) 194
hierarchical organizations 257–9
holism, *vs.* non-atomistic MI 142
Hollywood filmmakers 345–8
Homans, George C. 83
homophily 35, 308, 435
  choice 309–10
  induced 309, 310
  in network formation 310
  sources of 309
  variations of 19–20
households 457
  locally interdependent 458
Human Rights Campaign (HRC) 323–4
Hume, David 237, 394
hypothetico-deductive method of explanation 68

identification strategies 371–7
  collider variable 373–5
  conditioning 372–3
  front door criterion 375–6
  instrumental variables 376–7
Identities, Boundaries, and Social Ties (Tilly) 12
incidental agency 265, 281
incremental divergence 287–9
Independence of Irrelevant Alternatives (IIA) 192, 194
indeterminacy 274–7
  interactions and 276–7
indeterminacy, modes of 265
individual causal effect (ICE) 367, 368
individualism
  methodological, *see* methodological individualism (MI)
  structural, *see* structural individualism
individual-level variables 112–14
individual rationality 143–4
induced homophily 309, 310
inequalities
  educational 122
  generation of 19
  models 485–7
inequality of educational opportunity (IEO) 91
inequality of social opportunity (ISO) 91
influence-response functions 205, 208–13
  alternative 211–12
  repulsive 212–13
influencers 325–6
influences, countervailing 324–5
informal norms 234
institutional norms 234
instrumentalism 32
instrumental variables (IV) 376–7
intellectual risk 158
intentionalism of analytical sociology 156–7
intentionality, determinism *vs.* 155–6
interactions, *see* social interaction
interdependence 60–1, 279, 456
  causal 60n8
  strategic 60n8
interest rates 287–9
inter-individual divergence 287
internal validity 400–2
International Network of Analytical Sociology (INAS) 1, 8, 158
International Sociological Association 119
Internet Movie Database (IMDb) 337, 338, 339–40, 345–7
interpretativism 137
Interuniversity Center for Social Science Theory and Methodology (ICS) 83
intra-individual change 287
intrinsic value 259–60

James, William 69

Kahneman, Daniel 194
kitchen sink fallacy 331

lab experiments 392
language 67
Language, Truth, and Logic (Ayer) 67
law of large numbers 399
legal norms 234
legitimacy 321
Lindenberg, Siegwart 82–3
Lindenberg's model 87
Little, Daniel 64, 73, 74
localism, methodological 172–5
Logical Structure of the World (Carnap) 68
logic of appropriateness 260

machine learning 355
macro-level entities, ontological status of 144–5
macro-level outcomes 455
macro-narratives 157
macro-sociological theory of social structure 35
macro-sociology 13, 16
Maginot line 166
Making Sense of Marx (Elster) 74
manifestos, analytical sociology 2, 4, 7–8
manipulability theories of causation 394
Marx, Karl 73, *see also* analytical Marxism
Mathematical Sociology (MASO) 83
McClelland, David 50n1
Mead, George Herbert 69
meaning-making mechanisms 166

Gianluca Manzo - 9781789906851

mediation analysis, advances in 364–88
approach to 379–81
causal mediation effect 383–4
directed acyclic graphs 364, 369–71
identification strategies 371–7
mediating mechanisms 365–6
potential outcome model 366–70
Merton, Robert King 8n4
Merton award 1, 8–10
action theory 32–5
"anomalous" elements 35–8
articles' methods, data, and selected aspects of theoretical framework 23–4
data and methods 25–31
explananda 13–18
explanans 14–15, 18–22, 34
framing choices 11–13
modus operandi 31, 39–40
naming 8n4
papers winners of 159
signs of visibility 10–11
methodological individualism (MI) 85–8
criticisms of 139–41
non-atomistic 142–4
rejecting 139–41
social atomism 139
voluntarism 140
methodological localism 172–5
methodological statements 269–71
micro-level theories, structural individualism 88–9
micro-macro approach 80, 115–16
Micromotives and macrobehaviours (Schelling) 12
middle range theories 101, 114
reconstruction of 91–2
Mill, J.S. 220
model error 479
model of agent behavior 455, 457
Model of Frame Selection (MFS) 197–9
model of racial segregation 81
models
complexity 478–9
diversity 475–7, 483–5
inequalities 485–7
variety 475–7, 479–83
modernization 21
cultural 161
Moore, G.E. 67, 68
moral norms *vs.* social norms 235
motivation, to use Coleman diagram 53–4
move-unhappy-households 458
multi-mechanistic network model 435
mundane *vs.* experimental realism 401–2
Muslim women 21, 28

naïve realism 136–7
natural experiments 392–3
negative heuristics 5
network amplification 308–9
formation of network 309–11
of inter-group differences 312–14
mechanisms for network effects 315–16
network effects on behaviors and outcomes 311–12
network bias 329–30
network data 437
network effect 315–16
network mechanisms 432, 433–6
structural position of 433
network models 432–3, 436–7
empirical examples 437–8
for simulation 448–50
statistical, *see* statistical network models
network positions 433–4
networks
amplification, *see* network amplification
centrality to hierarchies 255–6
formation 309–11
networks of networks (NoN) 227
Neurath, Otto 68
Newton, Isaac 349
nodes
central 256–7
Coleman diagram 50–1
nominalist usage of contingency 268–9
non-atomistic methodological individualism 142–4
noncompliance 404
non-foundationalism of analytical sociology 138–9
normative ethics 68
normative influence 315
norms
coalition, group pressure as 244–7
cultural 241–4
formal 234
function of 234–5
group 233, 244–7
informal 234
legal/institutional 234, 238
moral norms *vs.* 235
overview 233
reciprocity 240
social 234
social dilemmas and 238
symmetrical cooperation 236–9
types of 234–6

Gianluca Manzo - 9781789906851

observational data 356
Observed Means Differences 397
One Deviation (OD) 199
online experiments 393, 407–11
  as field experiments 410–11
  pros and cons of 409
ontological claims 271–2
ontology of critical realism 135–8
open-ended conjunctures 274–7
open systems 137
Opp, Karl-Dieter 82
opportunity-based interactions 315
organizational demography models 252
orthodox decision theory 194–5
The Oxford Handbook of Analytical Sociology (Hedström and Bearman) 80

parallel-group, form of RCT 394
Parmenides (Socrates) 220
parsimony 187n2
path length 229–30, 322
  complex 326–8
  defined 327
Peirce, Charles 69
Philosophical Investigations (Wittgenstein) 67
Philosophy and Public Affairs (journal) 73
philosophy of psychology/neuroscience 68
philosophy of science 68, 70
  and structural individualism 84–5
philosophy of social science 72
political philosophy 68
popularity 434–5
positions, properties of 255–61
positive heuristics 5
positivism
  limitations of 135–6
  and naïve realism 136–7
postulates, synthetic set of 5
potential outcome model 366–70
pragmatism 170, 495
  action theory 175–8
  causality 178–9
  focus of 181–2
  methodological localism 172–5
  methods 179–80
  points of alignment 170–2
pragmatism, American 69
predictive policing 354, 354n2
pre-exposure prophylaxis (PrEP) 366
preference-constraints model (PCM) 195–6
  rationality and 189–91
preferential attachment 434–5
pre-formed agents 166
Principia Ethica (Moore) 67–8
Principia Mathematica (Newton) 349
Principia Mathematica (Whitehead and Russell) 66
Prisoner Dilemma (PD) 415–18
probabilities 191–2
Production of Commodities by Means of Commodities (Sraffa) 74
Protestant religious doctrine 55–6
Putnam, Hilary 69
puzzle 321
pyramidal impact 281

quadratic assignment procedure (QAP) regression 440–1
quasi-experiments 396
Quine, W.V.O. 69

racial residential segregation 20
racial segregation 81, 474–5
random assignment 396–9
random graph model 222–3
random interactions 265, 282
randomized control trial (RCT) 393–4, *see also* experiments
randomness 273, 298–302
random sampling 111–12
  disadvantage of 111–12
rational-choice Marxism 74
rational choice model (RCM) 141, 143, 160
rational-choice theory (RCT) 32, 34, 57–8, 127–31, 186, 189–93, 249–53
  adoption of 128–9
  behavioral assumptions of 130
  bounded rationality 195–6
  dual-process perspective 196–200
  paradigmatic privilege of 253
  rationality and the preference-constraints model 189–91
  realism of 128
  sociological 192–3, 194–5
  sociological applications of 128
  utilities and probabilities 191–2
  weaknesses of 253
rationality 33, 128–9
  bounded 195–6
  individual 143–4
  and preference-constraints model 189–91
  social, coupled contexts and 253–5
Rawls, John 69
A Realist Theory of Science (Bhaskar) 138n3
realist usage of contingency 268–9
reciprocity 434
reciprocity norms 240
reconstruction of middle range theories 91–2
reference, of noun 67
re-formulation of analytical sociology 5–6

Gianluca Manzo - 9781789906851

regularity theories of causation 394, 395
reinforcement learning 355
relational data 353–4
relational event models 446–8
religious doctrine 55–6
relocation cascade 456
Rényi, Alred 222
representative-agent approach 101, 103, 103n4, 110
repulsive influence-response function 212–13
Research Committee 28 (RC28) 119, 119n1
residential segregation 20, 27, 103–5, 115–16
  macro-phenomenon of 111
  unintended, ABCM of 456–69
restricted sample 405
retrodiction 150–1
retroduction 148–50
revised Kantianinsm 138
Robinson, John P. 166
Roemer, John 64, 73, 74, 77
Rorty, Richard 69
R-squared 479
RUSCORP database 337
Russell, Bertrand 66
Ryle, Gilbert 67

Sahlins, Marshall 273
same-ethnic friendship 22
same-norm inhibition effect 21
sanctions 237–9
sanctions, categories of 238
scale-free networks 224
Script Obedience 199–200
secularization 161
segregation model (Schelling) 103–5, 115–16, 187
selection bias 405
self-binding 161–2
self-selection 297–8, 337
semantic correlates 266–7
semiotic conception of culture 177
sense, of noun 67
sensor data 353
sequential equilibrium in game theory 194
sequential impact 281–2
sexually transmitted infections (STI) risk 373–4
Sierpiñski triangle 106–7
similarity, power of 330–1
simple contagion *vs.* complex contagion 322
simulation
  network models for 448–50
SIR model 480–3
situational context 458
situational mechanism 18, 86
small-world networks 224
small world network theory 229–30
Smith, Adam 241
social atomism 139, 142
social contagion 315
  models 163
social coordination 321
social dilemmas 236–9
  game-theoretic models of, *see* game-theoretic models
social dynamics 140
social environment 204
social identity theory 36–7
social inequalities
  methodological toolbox for describing 123–7
  network amplification of 315–16
  testing explanations of 131–2
  theoretical toolbox for explaining 127–31
social influence network theory 316n2
social interaction 204–6
  and indeterminacy 276–7
  influence-response functions 205, 208–13
  literature on 205
  macro-phenomena 205
  processes causing 206–7
  random 265, 282
  structure of 500–2
social learning 315
Social Mechanisms: An Analytical Approach to Social Theory (Hedström and Swedberg) 80, 490
social mechanisms and causality 144–7
social network 308, 432–3
  amplification, *see* network amplification
  cause and effect 228–30
  focus of empirical research 92–3
  positions and, properties of 255–61
  properties of 255–61
  research on 81
  segmentation 21
  social structure and 227–8
social network analysis (SNA) 436
social norms 234, *see* norms
social ontology 73
social pressure 207
social-psychological research 207
social rationality
  coupled contexts and 253–5
social roles 235
social sciences 148–51
social stratification
  multidimensional models of 260
Social Stratification Research (SSR) 119–20, 121, 122–3, *see also* social inequalities

Gianluca Manzo - 9781789906851

social structure 220
analytical sociology, perspectives in 225
defined 221, 222
philosophy, perspectives in 220–2
social and network sciences, perspectives from 222–5
and social networks 227–8
social tipping points 328–9
socio-cultural processes 157
sociological research 110–11
Sociological SEU theory 192–3
sociology 100–1
micro-macro problem in 101–5
shared research program for 120–3
Sociology as Population Science (Goldthorpe) 120
Socrates 220
Sørensen model 250–2, 250n3
spatial data 353
spontaneous stratification 107–8
spurious relationships 395, 395n1
statistical network models
conditional uniform graphs 438–40
Exponential Random Graph Models 441–3
quadratic assignment procedure regression 440–1
relational event models 446–8
stochastic actor-oriented models 443–6
statistical validity 401
status hierarchies 22
stochastic actor-oriented models (SAOM) 31
stochastic actor-oriented models (SAOMs) 443–6, 449
Stokman, Frans 82
strategic interdependence 60n8
stratification system, features characterizing 250–1
structural change 60
Structural Holes (Burt) 223
structural individualism 81–2, 173
as analytical sociology 94–6
developments since 1980s 93–4
features of 84–94
focus of empirical research 92–3
heuristics for micro-macro modeling 89–91
methodological individualism 85–8
micro-level theories 88–9
overview of 82–3
philosophy of science and 84–5
reconstruction of middle range theories 91–2
structural-individualistic research program 102, 103–4, 109
structure, *see* social structure
The Structure of Objects (Koslicki) 221
stylized network models 438–40
subadditivity 485
substantive statements 269–71
Sullivan, Oriel 166
supervised learning 355
Sure Thing Principle 194
survey-based research 123–7
SUTVA (Stable Unit Treatment Value Assumption) 368
Swidler, Ann 156
symmetrical cooperation norms 236–9
symptomatic contingency 271
system-imposed constraints 293–5

tangible archives 341–5
temporal data 353
textual data 353
theory of action 57–8
Theory of Planned Behavior (Ajzen) 193
Theory of Revealed Preferences (Kreps) 190
thick descriptions 155
thin descriptions 155
The Tipping Point (Gladwell) 321
tipping points 328–9
Tractatus Logico-Philosophicus (Wittgenstein) 67
tragedy of the commons 237
transcendental argumentation 147, 149
transcendental deduction 137
transfactual conditions 147
transformational mechanisms 86
transformational model of social explanation 141
transitive power relationships 257
transitivity 434
transparency 131
trust dilemma (TD) 416–17, 427–8
Tversky, Amos 194
typologies of social mechanisms 18–22

unstructured data 337, 337n2
unsupervised learning 355
update-households 458
utility 33, 33n8
utility functions 191

validity 399–400
ecological 403–4
external 402–3
internal 400–2
threats to 404–5
trade-offs between different forms 405–6
valuation error 479
Vanberg, Viktor 82
variable-based thinking 112–14

Gianluca Manzo - 9781789906851

variety, models 475–7, 479–80
in SIR model 480–3
veiling, and self-binding 161–2
"Venturi tubes" 252
Verstehen tradition 138, 141
video data 353
voluntarism 140
volunteer's dilemma (VOD) 420–1

Waismann, Friedrich 68
WARP with Limited Attention (WLA) 196
Watts, Duncan 230
Weak Axiom of Revealed Preference (WARP) 190–1, 192, 196
Weber, Max 55
Whitehead, Alfred North 66
Wippler, Reinhard 82
within-subjects design 407
women
Muslim 21, 28
underrepresentation in specialized jobs 19
Wood, Allen 73
Wright, Erik Olin 74

Gianluca Manzo - 9781789906851

Gianluca Manzo - 9781789906851