

Edited by

B. Guy Peters *and*
Guillaume Fontaine

EE
Elgar



*Handbook of Research Methods
and Applications in*

COMPARATIVE POLICY ANALYSIS

HANDBOOK OF RESEARCH METHODS AND
APPLICATIONS IN COMPARATIVE POLICY ANALYSIS

HANDBOOKS OF RESEARCH METHODS AND APPLICATIONS

Series Editor: Mark Casson, *University of Reading, UK*

The objective of this series is to provide definitive overviews of research methods in important fields of social science, including economics, business, finance and policy studies. The aim is to produce prestigious high quality works of lasting significance. Each *Handbook* consists of original contributions by leading authorities, selected by an editor who is a recognised leader in the field. The emphasis is on the practical application of research methods to both quantitative and qualitative evidence. The *Handbooks* will assist practising researchers in generating robust research findings that policy-makers can use with confidence.

While the *Handbooks* will engage with general issues of research methodology, their primary focus will be on the practical issues concerned with identifying and using suitable sources of data or evidence, and interpreting source material using best-practice techniques. They will review the main theories that have been used in applied research, or could be used for such research. While reference may be made to conceptual issues and to abstract theories in the course of such reviews, the emphasis will be firmly on real-world applications.

Titles in the series include:

Handbook of Research Methods and Applications in Entrepreneurship and Small Business

Edited by Alan Carsrud and Malin Brännback

Handbook of Research Methods and Applications in Spatially Integrated Social Science

Edited by Robert J. Stimson

Handbook of Research Methods and Applications in Economic Geography

Edited by Charlie Karlsson, Martin Andersson and Therese Norman

Handbook of Research Methods and Applications in Environmental Studies

Edited by Matthias Ruth

Handbook of Research Methods and Applications in Social Capital

Edited by Yaojun Li

Handbook of Research Methods and Applications in Transport Economics and Policy

Edited by Chris Nash

Handbook of Research Methods and Applications in Heterodox Economics

Edited by Frederic S. Lee and Bruce Cronin

Handbook of Research Methods and Applications in Happiness and Quality of Life

Edited by Luigino Bruni and Pier Luigi Porta

Handbook of Research Methods and Applications in Political Science

Edited by Hans Keman and Jaap J. Woldendorp

Handbook of Research Methods and Applications in Experimental Economics

Edited by Arthur Schram and Aljaž Ule

Handbook of Research Methods and Applications in Comparative Policy Analysis

Edited by B. Guy Peters and Guillaume Fontaine

Handbook of Research Methods and Applications in Comparative Policy Analysis

Edited by

B. Guy Peters

Maurice Falk Professor of Government, University of Pittsburgh, USA

Guillaume Fontaine

Professor, Department of Public Affairs, FLACSO, Ecuador

HANDBOOKS OF RESEARCH METHODS AND APPLICATIONS

 **Edward Elgar**
PUBLISHING

Cheltenham, UK • Northampton, MA, USA

© B. Guy Peters and Guillaume Fontaine 2020

Cover image: Miguel Sousa 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: 2019956360

This book is available electronically in the **Elgaronline**
Social and Political Science subject collection
DOI 10.4337/9781788111195

ISBN 978 1 78811 118 8 (cased)
ISBN 978 1 78811 119 5 (eBook)

Contents

<i>List of contributors</i>		xi
1	Introduction to the <i>Handbook of Research Methods and Applications in Comparative Policy Analysis</i> <i>B. Guy Peters and Guillaume Fontaine</i>	1
1	Introduction: substance and methods in comparative policy analysis	1
2	Theoretical issues	3
3	Methodological issues	6
4	Research design and methods selection	9
5	Conclusion: the importance of method for comparative policy research and practice	14
PART I THE METHODOLOGICAL DEBATE		
2	The comparative method and comparative policy analysis <i>B. Guy Peters</i>	20
1	Introduction	20
2	Modes of analysis in comparative policy studies	21
3	Sampling on the dependent variable	25
4	The comparative method and case studies	26
5	The possibility of hybrid designs	26
6	Several thoughts on small-N research	28
7	Conclusion: what, if anything, is special about comparative policy studies?	28
3	The most-similar and most-different systems design in comparative policy analysis <i>Carsten Anckar</i>	33
1	Introduction	33
2	The most-similar systems design	33
3	The most-different systems design	34
4	Deductive and inductive research strategies	35
5	Applying the most-similar systems design in deductive and inductive research	36
6	Applying the most-different systems design in deductive and inductive research	37
7	The most-similar and most-different systems designs compared	38
8	Criticism and response	39

9	The relevance of the most-similar and most-different systems in contemporary research	42
10	Conclusion	45
4	Can a case study test a theory? Types and tokens in comparative policy analysis <i>Keith Dowding</i>	49
1	Introduction	49
2	What is a theory?	49
3	Testing a theory as an invariant generalization	51
4	Testing a theory as an empirical generalization	54
5	Testing a theory as a mechanism	56
6	Unique cases	60
7	Conclusions	62
PART II THEORETICAL CHALLENGES		
5	Comparing policy processes: insights and lessons from the Advocacy Coalition Framework research program <i>Daniel Nohrstedt, Christopher M. Weible, Karin Ingold and Adam D. Henry</i>	67
1	Introduction	67
2	Elements of comparison when applying the ACF	68
3	Comparative ACF research: guidelines, experiences and shortcomings	70
4	Insights and lessons	74
5	Conclusion	81
6	Comparing agenda-settings: the Comparative Agendas Project <i>Laura Chaqués Bonafont, Christoffer Green-Pedersen and Henrik Bech Seeberg</i>	90
1	Introduction	90
2	Agenda-setting models	90
3	Methodological developments	92
4	Political agendas	95
5	Measures of the agenda	97
6	Future direction of policy agenda research	101
7	Comparing historical cases: advances in comparative historical research <i>Grace Jaramillo</i>	113
1	Introduction	113
2	Comparative historical analysis	114
3	Maximizing process tracing results	115
4	Indexes as tools of organizing and validating results	117
5	Case study: assessing industrial policy using an index of intervention	119
6	Conclusions	130

8	Comparing international policy transfers <i>Osmany Porto de Oliveira</i>	134
1	Introduction	134
2	The comparative analysis of policy transfers: a challenge, solution, and opportunity	136
3	Beyond comparative country analysis: transnational comparison	137
4	Issues, questions, and concepts for policy transfers comparative analysis	138
5	What and how should we compare in international policy transfers analysis?	140
6	Comparing transfer agents, instruments, and arenas	144
7	Conclusion	148
PART III MEASUREMENT AND EXPERIMENTAL METHODS		
9	Using experiments in comparative policy analysis: from policy evaluation to the policy process <i>Peter John</i>	153
1	Introduction	153
2	What are experiments?	154
3	Causal inference	155
4	The history of experiments in public policy	156
5	Elite experiments	159
6	Conclusions: lessons for students of the policy process	162
10	Measuring change in comparative policy analysis: concepts and empirical approaches <i>Jale Tosun and Julia Schnepf</i>	167
1	Introduction	167
2	Conceptualizing policy change	168
3	The multi-dimensional approach to measuring policy change	170
4	The budget approach to measuring policy change	175
5	The diffusion approach to measuring policy change	178
6	The outcome approach to measuring policy change	180
7	Discussion and conclusion	182
11	Using indexes in comparative policy analysis: global comparisons <i>Tero Erkkilä</i>	186
1	Introduction	186
2	Background	187
3	Governance indicators and critique of ranking	189
4	Towards second-generation measurements: transparency metrics	190
5	Measurement issues in regional and city-level indicators	193
6	Summary	198

7	Conclusions	199
12	Using text-as-data methods in comparative policy analysis <i>Fabrizio Gilardi and Bruno Wüest</i>	203
1	Introduction	203
2	Text-as-data applications in comparative policy analysis: an overview	204
3	Text-as-data in practice	207
4	Methods	210
5	New directions of text-as-data applications	212
6	Conclusion	213
PART IV MIXED METHODS AND MULTI-METHODS		
13	Critical multiplism for comparative policy analysis <i>William N. Dunn and B. Guy Peters</i>	219
1	Introduction	219
2	Critical multiplism and mixed methods	220
3	Some mixed-methods designs	225
4	Bounding multiple triangulation	232
5	Conclusions	233
14	Causal case studies for comparative policy analysis <i>Derek Beach</i>	238
1	Introduction	238
2	Key terms	239
3	Variance-based versus case-based approaches to studying public policies	240
4	How to do case-based studies of policy processes	245
5	Conclusions	250
15	Qualitative Comparative Analysis for comparative policy analysis <i>Eva Thomann</i>	254
1	Introduction	254
2	QCA as a technique	254
3	QCA in comparative public policy analysis	256
4	Approaches to QCA in comparative public policy	260
5	A case-oriented or condition-oriented approach to cases	261
6	A theory-generating or theory-evaluating mode of reasoning	263
7	Emphasizing substantively interpretable or redundancy-free models	265
8	Conclusions	270
16	Process tracing for comparative policy analysis: a realist approach <i>Guillaume Fontaine</i>	277
1	Introduction: a popular method	277

2	Methodology	279
3	A template for realist process tracing	283
4	Conclusion: process tracing for policy learning	291
PART V QUALITATIVE TECHNIQUES		
17	Using focus groups in comparative policy analysis <i>Patrik Marier, Daniel Dickson and Anne-Sophie Dubé</i>	297
1	Introduction	297
2	Why use focus groups in comparative public policy?	298
3	The deployment of focus groups in Québec	300
4	Methodological advantages	301
5	Methodological disadvantages	305
6	Designing focus groups	305
7	Managing group dynamics	307
8	Conclusion	308
18	Using ethnography in comparative policy analysis: premises, promises and perils <i>Raul Pacheco-Vega</i>	312
1	Introduction	312
2	What is ethnography and how can it be applied in comparative policy analysis?	313
3	Ethnography in public policy, public management, and public administration research	315
4	Three modes of ethnographic engagement in comparative policy analysis	317
5	Comparative policy analysis and policy studies from a comparative perspective: the two faces of Janus?	321
6	Which policy issue areas use ethnography?	323
7	Doing ethnography for comparative policy analysis: an applied approach	323
8	Promises and perils of the application of ethnography as a research method for comparative policy analysis	326
9	Conclusion: embracing ethnography in comparative policy analysis?	328
19	Using Q methodology in comparative policy analysis <i>Astrid Molenveld</i>	333
1	Introduction	333
2	Philosophy	334
3	How to execute and analyze a Q-methodology study	335
4	Conclusion	343
20	Using the Narrative Policy Framework in comparative policy analysis <i>Aaron Smith-Walter and Michael D. Jones</i>	348

1	Introduction	348
2	The narrative policy framework	348
3	Universality of narrative	351
4	Traveling with the NPF	355
5	Limitations	360
6	Conclusion	360
PART VI ISSUES FOR FURTHER RESEARCH		
21	Trends in the development of comparative policy analysis <i>Iris Geva-May, David C. Hoffman and Joselyn Muhleisen</i>	367
1	Introduction	367
2	Comparative policy analysis development as a distinct field of study	368
3	The importance and limitations of comparative policy analysis and comparative policy analytic studies	371
4	Trends in comparative policy analysis refereed articles: content analysis of EBSCO all academic database and the JCPA 1998–2016	374
5	The <i>Journal of Comparative Policy Analysis</i>	378
6	Conclusions	379
22	Evolutionary theory in comparative policy analysis <i>Adrian Kay</i>	385
1	Introduction	385
2	Evolutionary approaches	386
3	Evolutionary frameworks for institutionalist theory in comparative policy analysis	392
4	Evolutionary theory in comparative public policy: methodological issues	394
5	Conclusion	396
	<i>Index</i>	401

Contributors

Carsten Anckar is Professor of Political Science (Comparative Politics) at Åbo Akademi University, Finland.

Derek Beach is Professor of Political Science at Aarhus University. His research focuses on elite decision making, policy learning, and case study methodology.

Laura Chaqués Bonafont is Professor of Political Science at the University of Barcelona and research fellow at the Institut Barcelona d'Estudis Internacionals (IBEI). She is leading the research group Quality of Democracy (www.Q-Dem.com) which aims to analyse and create data to better understand agenda dynamics in comparative perspective.

Daniel Dickson is a PhD candidate in Political Science at Concordia University. His dissertation project is the development of a cross-provincial policy index to compare effectiveness in implementing social inclusion policies for older Canadians with developmental disabilities. Drawing from his past experience in the social services sector, his primary research interest is in analysing the effects of public policies designed to reduce social marginalization and exclusion, with a specific focus on disability and population ageing.

Keith Dowding is Distinguished Professor of Political Science and Political Philosophy at the Australian National University, Canberra. He has published over 20 books including *The Philosophy and Methods of Political Science* (Palgrave Macmillan, 2016).

Anne-Sophie Dubé currently works for the Centre de recherche et d'expertise en gérontologie sociale (CREGÉS) in Montreal, Québec. She has a Master's degree in public health (specialization in epidemiology) at Université de Montréal. Her primary research interests include population health, active transportation and the built environment.

William N. Dunn is Professor of Public Policy at the University of Pittsburgh. He is author of *Public Policy Analysis: An Integrated Approach*, 6th ed. (Routledge, 2018) and *Pragmatism and the Origins of the Policy Sciences: Rediscovering Lasswell and the Chicago School* (Cambridge University Press, 2019). His awards include the Donald T. Campbell, Aaron Wildavsky, and Douglas M. McGregor awards and the University of Pittsburgh Provost award for Excellence in Mentoring. He is an elected fellow of the National Academy of Public Administration.

Tero Erkkilä is Associate Professor of Political Science at the University of Helsinki. His research interests include global knowledge governance, public institutions and collective identities.

Guillaume Fontaine is senior researcher at the Latin American Faculty for Social Sciences (FLACSO) in Ecuador. He has published extensively on comparative policy analysis, democratic governance and policy design in Latin America, including *The Politics of Public Accountability* (with Cecilia Medrano and Ivan Narváez, Palgrave Macmillan, 2019).

Iris Geva-May has been recognized by Thomson Reuters for having pioneered the field of comparative policy analysis since 1998, when she founded the highly cited *Journal of Comparative Policy Analysis* in 1998, and the International Comparative Policy Analysis Forum Scholarly Society. She is the author of numerous articles and books among which *Thinking Like a Policy Analyst: Policy Analysis as a Clinical Profession* (Palgrave Macmillan, 2005) and *An Operational Approach to Policy Analysis: The Craft* (Springer, 1997) and editor of *Classics of Comparative Policy Analysis* (Routledge, 2019), International Library of Comparative Policy Analysis Series (Policy Press), and *Routledge Handbook of Comparative Policy Analysis* (Routledge, 2017). She is currently a honorary visiting professor and Adjunct at Carleton University, Canada, the Wagner School NYU, and Professor Emerita, Vancouver, Canada.

Fabrizio Gilardi is Professor of Policy Analysis in the Department of Political Science at the University of Zurich. His research interests include policy diffusion, gender and politics, and digital technology and politics.

Christoffer Green-Pedersen is Professor of Political Science at Aarhus University. His research focuses on party competition, agenda-setting and public policy.

Adam D. Henry is an Associate Professor in the School of Government and Public Policy at the University of Arizona, USA, and is affiliated with the University of Arizona Institute of the Environment. His research applies network science and computational social science to the study of learning in the policy process.

David C. Hoffman earned his PhD from the University of Iowa in 2000 and has served on the faculty of the Marxe School of Public and International Affairs since 2003, where he is an Associate Professor. He has published work on classical rhetorical theory, as well as rhetorical criticism of political figures, both historical and contemporary, and has interests in deliberation and computer-aided textual analysis.

Karin Ingold is an Associate Professor of Policy Studies and Environmental Governance at the Institute of Political Science and the Oeschger Center for Climate Change Research, University of Bern, Switzerland. She also holds an affiliation to the Water Research Institute, Eawag, of the ETH domain. She is interested in research about complex decision-making processes applying the Advocacy Coalition Framework and tools borrowed from Social Network Analysis.

Grace Jaramillo, PhD, is Postdoctoral Fellow and lecturer at The University of British Columbia, Department of Political Science. She is a former professor at the Latin American Faculty for Social Sciences, FLACSO, where she became the chair of the Department of International Relations from 2007 to 2010.

Peter John is Professor of Public Policy at King's College London. He is known for his work on the theory of public policy and citizen engagement. His most recent books are: *How Far to Nudge* (Edward Elgar, 2018) and *Nudge, Nudge, Think, Think* (Bloomsbury, 2019).

Michael D. Jones is an Associate Professor at Oregon State University's School of Public Policy and Editor-in-Chief of the *Policy Studies Journal*. His research interests include the Narrative Policy Framework and policy process theory. His recent research has appeared

in top journals such as *Political Psychology*, *Critical Policy Studies*, and *Social Science Quarterly*, among others.

Adrian Kay is Professor of Politics and Public Policy at Swansea University in the UK. He has previously held positions in Australia, Asia and Europe. Adrian works at the intersection of international and comparative public policy, with a particular interest in Islamic governance.

Patrik Marier is a Professor of Political Science at Concordia University, where he holds a Concordia University Research Chair in Aging and Public Policy. He is also the Scientific Director of CREGÉS, a university research centre housed at the Integrated Health and Social Services University Network for West-Central Montréal. Dr Marier's research focuses on the impact of ageing populations on a number of public policy fields including pensions, health-care, and labour policy across comparative cases.

Astrid Molenveld is Assistant Professor at the Department of Public Administration and Sociology of the Erasmus University Rotterdam. She has a particular interest in applying multiple research methods in her work, like QCA, Q-methodology and statistics. Her current research activities include comparative research on coordination and collaborative governance, with a special focus on cross-cutting and 'wicked' policy issues.

Joselyn Muhleisen is a Doctoral Lecturer at the Marxe School of Public and International Affairs, where she teaches courses on regulatory policy, comparative public policy and public administration, and global economic governance. She earned her doctorate in political science from The Graduate Center, CUNY. She has published work about the development of comparative policy analysis and its relationship to international studies.

Daniel Nohrstedt is an Associate Professor of Political Science in the Department of Government at Uppsala University, Sweden, and research coordinator in the Centre of Natural Hazards and Disaster Science (CNDS). His research interests include public policy and the policy process, particularly through the lens of the Advocacy Coalition Framework.

Raul Pacheco-Vega is an Assistant Professor in the Public Administration Division of the Centre for Economic Teaching and Research (Centro de Investigacion y Docencia Economicas, CIDE) in Mexico. He is a specialist in comparative public policy and focuses on North American environmental politics, primarily sanitation and water governance, solid waste management, neoinstitutional theory, transnational environmental social movements and experimental methods in public policy.

B. Guy Peters is Maurice Falk Professor of Government at the University of Pittsburgh and was founding president of the International Public Policy Association. He is currently editor of the *International Public Policy Review* and Honorary Editor of the *Journal of Comparative Policy Analysis*.

Osmany Porto de Oliveira is Assistant Professor of International Relations at the Federal University of São Paulo (Unifesp). He holds two PhDs in Political Science (University of Sorbonne Nouvelle – 2015 and University of São Paulo – 2013). He received his BA in International Relations from the University of Bologna (2006) and an MA in Sociétés Contemporaines: Europe – Amériques from the University of Sorbonne Nouvelle/IHEAL (2008). In 2019 he received the 'Early Research Award' of the International Public Policy Association (IPPA). His main books are *Latin America and Policy Diffusion: From Import*

to *Export* (Routledge, 2020), with Cecilia Osorio, Sergio Montero and Cristiane Leite, *International Policy Diffusion and Participatory Budgeting: Ambassadors of Participation, International Institutions and Transnational Networks* (Palgrave MacMillan, 2017) and *Le transfert d'un modèle de démocratie participative: Paradiplomatie entre Porto Alegre et Saint-Denis* (Éditions IHEAL/CREDA, 2010).

Julia Schnepf has studied political science, economics and psychology at Heidelberg University. At present, she is doing her PhD in Political Psychology at the University of Koblenz-Landau. Her research addresses explanations of non-voting and extreme voting decisions, party polarization and populism, and methodological questions on voting behaviour.

Henrik Bech Seeberg is an Associate Professor in the Department of Political Science at Aarhus University. His research focuses on political agenda-setting at the national and local level in addition to party competition and its policy consequences. His research appears in *European Journal of Political Research*, *West European Politics*, *Party Politics*, *Journal of European Public Policy*, *Policy Studies Journal*, and *Journal of Public Administration Research and Theory*.

Aaron Smith-Walter is an Assistant Professor of Political Science and core faculty member in the Master of Public Administration programme at the University of Massachusetts Lowell. His research interests include the narrative policy framework, Grid-Group Cultural Theory, and images of public administration in popular fiction. He serves as Vice-President of the Massachusetts Chapter of the American Society of Public Administration and is an Associate Editor of the *Policy Studies Journal*.

Eva Thomann from the University of Exeter specializes in public policy, public administration, and case-oriented and set-theoretic research methodology. Her research about or applying Qualitative Comparative Analysis has appeared in international journals including *Sociological Methods & Research*, *European Journal of Political Research*, *Policy Sciences*, and *Journal of Public Administration Research and Theory*, and she is working on a co-authored textbook about QCA in R for Cambridge University Press. Eva is a co-chair of the international COMPASSS methods network and teaches QCA and related techniques at various international method schools.

Jale Tosun is Professor of Political Science at Heidelberg University and a Co-Director of the Heidelberg Center for the Environment. Her research interests comprise various topics in comparative public policy, comparative politics, international political economy, public administration and European studies.

Christopher M. Weible is Professor with the School of Public Affairs at the University of Colorado, Denver, USA. He co-directs the Workshop on Policy Process Research (woppr.org).

Bruno Wüest is a data analyst at sotomo (<https://sotomo.ch/>) and former senior researcher at the Department of Political Science of the University of Zurich. His principal research interests are computational social science, political communication, and political behaviour.

1. Introduction to the *Handbook of Research Methods and Applications in Comparative Policy Analysis*

B. Guy Peters and Guillaume Fontaine

1. INTRODUCTION: SUBSTANCE AND METHODS IN COMPARATIVE POLICY ANALYSIS

Comparative policy analysis (hereafter CPA) is an emerging area in social sciences and already a major one in policy studies. As of December 31, 2018, the Web of Science's database registered 243 references of peer-reviewed articles, whose authors used "comparative policy" as descriptor (as key words, in the title or in the abstract). This is still less than "comparative politics" (776 references during the same period), but it is much more significant than "comparative public administration" (39 references), considering the same token. More important, though, is the upward trend of the literature explicitly referring to CPA as a descriptor of the substance and method of a research, as more than 56% of these titles were published after mid-2015.¹

Given the multi-disciplinary nature of policy studies, the methodological challenges in CPA are arguably more complex than for other research areas in social sciences. Therefore, there is a greater need to provide an overview of the available methods, as well as their strengths and weaknesses for addressing theoretical issues like policy change, time and context in policy design, and so on. Hence, this handbook of methods for comparative policy analysis is intended to advance the understanding of methodology in the study of comparative public policies, as much as to broaden the array of methods and techniques considered by CPA scholars in their research design. The methods discussed in its 22 chapters, including this introductory study, consider not only the nature of public policies but also how to deal with the complexities of policy across political systems and across policy domains and across time. These methods are also of interest for practitioners, who face many policy problems of international dimensions and have a lot to gain from policy learning and policy diffusion, but often lack the methodological tools to deal systematically with the comparison of most-similar and most-different systems.

The need for such a handbook has become especially pressing for two additional reasons. The first is the increasing diversity and complexity of available methods in social sciences, which makes their selection a challenging exercise for policy design in research and practice, beyond substantive issues. Some are of relatively recent vintage and have been adding significant analytic power to CPA, like the qualitative comparative analysis methods (QCA) (see **Chapter 15** by Thomann). Others have been common in some disciplines but are now becoming more and more popular among policy scholars and practitioners, like process tracing (see **Chapter 16** by Fontaine), focus groups (see **Chapter 17** by Marier, Dickson, and Dubé), and ethnography (see **Chapter 18** by Pacheco-Vega). Still others are actually decades old but have

recently experienced a new appraisal as outstanding quantitative and qualitative techniques in the social sciences, like text-as-data analysis (see **Chapter 12** by Gilardi and Wüest) and the Q-methodology (see **Chapter 19** by Molenveld).

The other compelling reason for considering the range of methods available to CPA scholars is the necessity to overcome the limits of quantitative techniques (especially standard linear regressions and econometric models), which are sometimes mistakenly identified as the best way (if not the only effective one) to conduct cross-case, cross-area, or longitudinal comparative analysis. The overwhelming publication of probability-driven research outputs in high-impact journals – sometimes assimilated to grounded theory or to knowledge-based policy making – may produce the wrong impression that contemporary social sciences, in general, and CPA, in particular, rely on a limited range of methods considered to be valid. Indeed, probabilistic methods are generally used with little consideration of available alternatives, because of the assumptions that are involved in the regression or model designed by scholars for theory-building or testing. Moreover, they are often used in a deterministic fashion, following an oft-cited motto in public management “if you can’t measure it, you can’t improve it”.

We consider this reliance on a limited range of methods unfortunate for three major reasons. First, it affects the research agenda in CPA by neglecting major theoretical and practical issues for policy analysis and design: focusing on a limited number of methods will also limit the range of issues to be taken into consideration. Yet not all research questions are amenable to the same methods (see our typology of methodologies below). Second, it is often desirable to have results using one method compared with those from an alternative method, in order to gain additional insight into a research question (Webb et al. 1999). Taken as a whole, this handbook demonstrates that mixed- and multi-methods can provide a more complete and nuanced account of causal relationships (see in particular **Chapter 13** by Dunn and Peters). Third, understanding policies and policy choices requires detailed examinations of those individual policies and their consequences, which often implies conducting case studies or small-N comparison (see **Chapter 14** by Beach).

A more open-minded and conscious selection of methods and techniques to address a research or a policy problem should improve the quality of CPA. Hence spreading our methodological net as wide as possible is likely to provide more robust insights. Based on these premises, this introductory study raises two central questions that run throughout the handbook and are attended differently by the authors: What are the best available methods to conduct systematic CPA across time, space, and areas? And how can scholars and practitioners select them or combine them to improve the internal consistency and the external coherence of policy design in research and practice?

The remainder of this chapter is divided in four sections. Section 2 presents four major theoretical issues currently under discussion in CPA, in order to show the methodological implications of theory-driven research and how these can be attended by the methods presented in this handbook. It considers the structure versus agency debate, the study of policy change and the role of context and time. Section 3 presents the methodological issues based on a sample of 80,000 peer-reviewed articles from the Web of Science. It describes the methodological shift in social sciences and the methods preferences by areas thereof. Section 4 draws on the research design and methods selection stemming from the former sections. It builds on a typological model to deal with the diversity and complexity of these methods, and then it proceeds with comparing the scope and limitations of extensive and intensive research designs

in outcome explanation or interpretation. Eventually this section underlines the difficulties raised by case selection in a within-case study or a small-N comparison. Section 5 draws on some provisional conclusions that stress the importance of methods for comparative policy research and practice and presents the handbook's organization.

2. THEORETICAL ISSUES

Many theoretical perspectives are relevant for CPA in one way or another (Peters and Zittoun 2016; Weible and Sabatier 2017). In a research design these theories need to be connected with methods that can enable scholars to confirm or to disconfirm their hypotheses. If CPA is to provide a cumulative knowledge to improve policy design in theory and practice, such methods need to be capable of testing them comparatively. In this section we address some of the central issues at stake in the contemporary debate, as related to their methodological implications.

2.1 Structure versus Agency

The question of how structure and agency are related in explaining outcomes is pervasive in social sciences in general, and in CPA in particular. How much does the institutional structure within which decisions are made matter, as compared to the influence of individuals who actually make these decisions? In contemporary political science and sociology this question is at the core of the debates between *methodological individualism*, favoring behavioral and rational choice theories, and *neo-institutionalism*, arguing for the causal determination of structure. The correct answer to the question, of course, is that both matter but that does not solve the problem of where scholars should begin their analysis.

The structure-versus-agency debate appears particularly important in policy studies, since a policy is essentially made within institutions (see **Chapter 22** by Kay). That being said, there is undoubtedly a danger in anthropomorphizing institutions, thereby ignoring the nature of the individuals who occupy positions within those institutions (Scharpf 1997). The bottom line is that individuals make decisions that result in policies being adopted and implemented, but they do not act in a vacuum. They are influenced by the institutions within which they work, and with which they interact. Therefore, we would assume (with good cause) that different institutions will process policies differently and will privilege certain types of policy choices.

Thus, thinking about only one of these approaches to explanation is likely to provide only a partial vision of what is occurring when the policy is made. As already noted, this theoretical problem is endemic in the social sciences but the importance of institutions in a policy area – legislatures, bureaucracies, interest groups, etc. – makes the problem more acute. The extent to which individual decisions and the norms and information held by institutions may be intertwined is difficult to untangle, and may well require qualitative methods to identify the premises on which decisions are made.

2.2 The Role of Context

The status of context in comparative politics is somehow paradoxical. On one hand, the methods used for comparative analysis sometimes attempt to eliminate those contextual

factors when focusing on the relationship between an independent and a dependent variable. In particular, the comparative method as described by Arend Lijphart attempts to hold constant as many contextual factors as possible in order to assess those relationships (see **Chapter 2** by Peters). On the other hand many theories, especially those derived from comparative politics, attempt to involve contextual factors directly in the analysis of policy, through most-similar or most-different systems research designs (see **Chapter 3** by Anckar).

This seemingly paradox points to the importance of within-case and small-N comparison in theory-testing (see **Chapter 4** by Dowding). Within-case analysis allows for a full exploration of the contextual factors that influence a policy area, and that detailed analysis is not “mere description” (Gerring 2010), but does enable a full understanding of the dynamics of policy. But the comparative method as framed by Lijphart is concerned with cross-case analysis, and developing generalizations from multiple cases. In that sort of analysis, the context – as important as it is – may produce spurious results and therefore needs to be controlled in the research design.

One of the virtues of CPA is that it brings context directly into the analysis. Policy making does not occur in a social, cultural, or economic vacuum: it is influenced – and perhaps decisively so – by environmental factors. This is true even when experimental designs for research are being utilized: the context of the experiment may still matter, and should be included in the conceptualization of the research project (see **Chapter 9** by John). Some methods discussed in this handbook emphasize culture and context when attempting to assess the narratives elaborated by a policy area’s actors (see **Chapter 20** by Smith-Walter and Jones) but those factors are also important for any study of policy. For example, there is a significant body of literature arguing that economic factors are much more important than political ones for determining public spending. It is crucial to understand such factors as political culture and social structure when attempting to understand policy choices, and the effects of those choices within a society.

2.3 The Study of Policy Change

To a great extent the study of contemporary public policy is the study of policy change, given that there are few virgin areas into which governments can intervene (Carter 2012; Hogwood and Peters 1987). Most policy domains are occupied by a host of policies, so policy making basically consists in changing those existing policies and only occasionally fitting some new policy into that crowded domain. Consequently one of the main methodological challenges for CPA is coping with change.

A major modality of policy change is diffusion and learning (see **Chapter 8** by Porto de Oliveira). One of the virtues of thinking comparatively about public policy is that it rapidly becomes apparent that the world is a laboratory for policy change, where individual policies are tantamount to natural experiments. Policy innovators and activists are at work in many countries attempting to develop better solutions for common problems, and those ideas then are diffused across countries. There is an extensive literature on learning and diffusion (Dunlop and Radaelli 2013; Rose 1993) that points to the importance of understanding how policies travel across political systems, as well as the difficulties in making those transfers in practice. This interest in diffusion and learning has increased with the concern for “evidence-based policy making” and the ability to utilize evidence from one setting to inform policy design in

others (Cairney 2016; Pawson 2006), despite the inherent difficulties in translating the success in one setting to another.

Among all the dimensions of policy change, the most obvious is that governments who decide to change existing policies have to proceed differently according to the political systems in which they are embedded. For example, there is an extensive literature on the welfare state retrenchment among the developed democracies that provides important examples of how to understand that kind of policy change (Castles 2004; Starke 2006; Van Kersbergen and Vis 2014). That change may be the result of political change, policy innovation, policy entrepreneurship, or even the whims of a powerful individual (agency). To study this process requires measuring the nature of the policy at one point in time (either by quantitative or qualitative means), measuring the extent to which that measurement has changed, and then attempting to understand the difference (see **Chapter 10** by Tosun and Schnepf).

Once again, CPA should be especially sensitive to the contextual issues involved in utilizing evidence from policy making across cultures and societies. Some indicators of a public policy (e.g. public expenditures) may be relatively simple and usable across cultural boundaries, but many others are embedded in their social and cultural environments (see **Chapter 11** by Erkkilä). This list of concerns about CPA could be extended, but the bottom line is that identifying and measuring the variables involved is not as easy as for voting behavior, or public opinion.

2.4 The Problem of Time

Stemming from the problem of policy change, time is of the essence in CPA. The effects of a policy may be spread across time, hence requiring some consideration of the long- and short-term consequences of action (see **Chapter 7** by Jaramillo). Although most explicitly assumed CPA regards cross-case analysis, within-case longitudinal comparison is an equally important form of analysis. One could, in fact, argue that it is more important because it holds constant many possible confounding factors – the context mentioned above – and therefore allows for a clearer understanding of the relationships among variables in the analysis.

Yet this raises a twofold problem: a methodological one and a theoretical one. On one hand, techniques such as time-series and pooled time-series attempt to identify processes of development and change within a policy and relate them to changes in the presumed independent variables. But as powerful as studying policy across time can be, it also raises significant methodological challenges (see Kellstedt and Whitten 2013). The most common statistical problems met in CPA to assess the necessity and sufficiency of a condition are related to omitted variables (which are virtually infinite in open systems) and multi-causality (i.e. competing theories of causation of a single phenomenon). Other issues require building complex econometric models, to deal with spuriousness (i.e. a variable affecting both the cause and the effect), equifinality (i.e. different independent variables causing a similar effect on a dependent variable), circularity (i.e. independent variables affecting one another), or heteroscedasticity (i.e. a variation among variables from the same set).

On the other hand, time does imply change in the context, and can be a significant source of invalidity in findings (Shedish, Cook, and Campbell 2002). Time, in a simple chronological sense, may be important because of external changes, but also because of changes within the subjects of the research (maturation) and changes in the researchers themselves (instrumentation). Dealing with these sources of invalidity is paramount both for qualitative

and quantitative research. Moreover, time-series analysis may explain very high correlations among factors in regressions, simply because the observations are not independent. There are statistical techniques to address this problem inherent to longitudinal comparison, but those statistical methods used to eliminate some of the effects of time may also undermine the understanding of some of the important effects of change across time. For qualitative researchers the effects of time will be more difficult to overcome.

Eventually, the difficulties associated with time in CPA may lead some scholars to take a more static approach to the phenomena they are interested in. But an assumption that variables, at any one point in time, do not have a past is likely to be fallacious. This produces another research question of where to draw the line when considering the influence of history. To understand contemporary policy in France, do we need to go back to World War II, to the Revolution, or to the *ancien régime*? The answer may depend on the particular policy question under study, and then again CPA scholars do need to understand the context. For what's left, understanding the background will illuminate the decisions being made at the present time, for any policy question.

In a nutshell, contemporary debates in CPA about structure and agency, policy change and the role of context and time call for different research designs, combining qualitative and quantitative methods. But how are scholars to combine these techniques in their research design? The following section provides some substantial elements based on bibliometrics to answer this question.

3. METHODOLOGICAL ISSUES

CPA shows some particular features that justify bigger methodological concerns than in other research domains from the social sciences. The most important one is that the field is at the confluence of various disciplines, which have been developing their own strategies of description, explanation, and interpretation. In this section we characterize what we call “the methodological shift” in social sciences and we picture how this transformation has spread across many different research areas which are of interest for CPA.

3.1 The Methodological Shift in Social Sciences

Like any other field of research, CPA must make fundamental choices about the most appropriate means of developing evidence. But in addition to the standard concerns with methods, whether coming from within disciplines or from the confluence of disciplines, CPA itself may pose additional challenges to the prevailing methodologies, the most important of which is the complexity of the subject matter being studied. This is why some analytical frameworks have been more successful than others in explaining policy choices, either focusing on the role of advocacy coalitions (see **Chapter 5** by Nohrsted, Weible, Ingold, and Henry) or long-standing trends in the agenda-setting process (see **Chapter 6** by Chaqués Bonafont, Green-Pedersen, and Seeberg).

The discussion concerning the proliferation of methods – hereinafter referred to as the methodological debate in CPA – partly reflects the contemporary concern with methods and techniques in social sciences. As of December 31, 2018 the Web of Science’s database registered almost 80,000 peer-reviewed articles claiming to use one or more methods described in

the present handbook.² This debate was sparked off during the heyday of comparative politics, political economy, and historical sociology, back in the 1970s. But the methodological shift actually goes back to the early 1990s, when the total number of articles assuming explicitly at least one of these research methods surpassed 100 per year. Another threshold was reached in 2002 with more than 1,000 articles published in a single year, and another one in 2017 with 10,000 articles (WOS 2019). It is noteworthy that more than half of these references were published between 2014 and 2018.

If we take political science, economics, and sociology as the “heartland” of CPA, then existing methodological discrepancies among scholars may not be too extreme. As a matter of fact, a majority of journals specialized in these three disciplines largely praise quantitative techniques (especially standard linear regressions) and experiments for theory-building and testing (Bardsley 2010). The ongoing debate, within these disciplines, between advocates of quantitative, qualitative, and multi-methods, reached a tipping point with the seminal book by King, Keohane, and Verba (1994). Yet, even if these authors and others (Brady and Collier 2010; Goertz and Mahoney 2012; Seawright 2016) state that the standards for quantitative and qualitative research are the same when it comes to theory-driven research, they undoubtedly show a bias towards a probabilistic approach to causation and causality.

This strategy has been challenged by scholars interested in institutional and cultural influences, arguably more in political science and sociology than in economics (Peters 2013). Some argue in favor of a configurational approach to causation through QCA, which has quite recently become popular in CPA (Engeli and Rothmayr Allison 2014; Ragin 2008; Rihoux and Grimm 2006). Others value the contribution of case study and small-N comparison as complementary qualitative methods (Beach and Pedersen 2016; Blatter and Haverland 2012). Still others advocate for set-theoretic research designs based on multi- or mixed methods (Berg-Schlosser 2013; Blatter and Haverland 2014).

That being said, those are by no means the only three contributing disciplines to CPA. Many others bring in complementary approaches, hence providing new insights on causality and causation. For instance, anthropology brings in a range of observational techniques – like focus groups, ethnography, discourse analysis, etc. – that rely more on the locus of the individual researcher than on existing quantitative indicators. Further, both law and history work more with documentary evidence through congruence analysis, process tracing, and other qualitative techniques. Last but not least, substantive policy domains – such as public health, international relations, environmental sciences, or even engineering – have made original contributions to the debate, to be integrated with those of the conventional approaches in social sciences. These alternative sources of evidence contribute to the development of CPA, even if they can also be a source of confusion and conflicting elements for the coherence and consistency of a research design, as will be discussed below.

3.2 Methods Preferences by Research Areas

The diversity of methods in social sciences and CPA makes it worth a thorough review of their historical trends before discussing their contribution to research and practice. Based on our sample of peer-reviewed articles, the literature can be divided into four groups. The first one includes research papers based on case study, small-N comparison, or experiments, three methods that are widespread in social sciences, for they have been utilized since the 1970s. The second one consists in the research papers grounded in quantitative methods such as

surveys, standard linear regressions, and time-series, which have been increasingly used since the early 1990s. The third one is made of qualitative methods and techniques whose utilization has been increasing since the early 1990s, including ethnography, focus groups, comparative historical analysis, multi- or mixed methods, and the Q-methodology. A fourth group includes papers based on methods and techniques that already existed in the 1980s but became standard operating procedures in CPA during the 2000s, such as the comparative method, QCA, and process tracing. Scholars may combine various methods; therefore the same article can appear in different groups.

All these methods actually benefited from two major events affecting policy making. On one hand they were fueled by the exponential effects of innovations in information and communication technologies, from the design of the first statistical software packages available to the general public in the 1980s, and from the growing industry of big data in the 2010s. On the other hand they experienced a faster and wider diffusion across the academy and the practitioners community due to the intensification of globalization effects, including the emergence and the multiplication of transnational policy problems, the digitalization of knowledge, and the blossoming global community of policy scholars embodied in the International Political Science Association (created by UNESCO in 1949) and its offspring, the International Public Policy Association (born in 2016), not to mention all regional associations like the European Consortium for Political Research (ECPR, born in the 1970s).

Cross-area data comparison indicates scholars' preferences for certain methods and techniques according to their research area (see Appendix table).

Group 1 represents 64.81% of the overall literature. Two-thirds of the research based on case study and small-N comparison are concentrated in six areas: environmental sciences and studies (28%), public administration and management (10%), economics (8%), political science (6%), education (6%), and geography (6%). More than half of the literature using experiment is to be found in five areas: economics (15%), environmental sciences and studies (13%), computer science (10%), political science (8%), and public administration and management (8%).

Group 2 represents 11.66% of the sample. It is also concentrated in a few substantive areas, as two-thirds of the research based on quantitative techniques and methods focus on three areas – economics (38%), environmental sciences and studies (18%), and public administration and management (14%). Scholars using these methods are also relatively interested in political science (9%) and international relations (3%).

Group 3 concentrates 22.25% of the total. Methods used in this group are more evenly distributed across areas. First, two-thirds of the research using ethnography are concentrated in six areas – anthropology (18%), sociology (11%), education (11%), geography (9%), international relations (8%), and public administration and management (7%) – but this research design is also applied to environmental studies and sciences (4%), political science (4%), and linguistics (4%). Second, two-thirds of the research based on comparative historical analysis are to be found in five areas – political science (17%), environmental sciences and studies (17%), economics (14%), public administration and management (10%), and geography (8%) – but this method is also utilized in sociology (6%), international relations (6%), and history (5%). Third, most research using the Q-methodology is utilized in eight areas – education (16%), environmental sciences and studies (16%), linguistic (12%), communication (11%), political science (9%), sociology (9%), geography (6%), and international relations (6%).

That being said, some qualitative techniques from group 3 are still overlooked in a variety of research areas. Focus groups techniques basically focus on four areas: public administration and management (22%), health care sciences, health policy, and general medicine (21%), environmental sciences and studies (11%), and education (9%). Two-thirds of the research based on multi-methods and mixed methods are concentrated in five areas – health care and health policy (18%), public administration and management (17%), environmental sciences and studies (13%), education (11%), and international relations (5%).

Group 4 represents 1.3% of the total. This one is also concentrated in a few areas but, as already mentioned, this is a group of emerging techniques that barely sums up 1,028 papers, which makes it less representative. The comparative method is concentrated in five areas – public administration and management (16%), environmental sciences and studies (13%), health care and health policy (13%), political science (11%), and international relations (5%). Process tracing is concentrated in political science (48%), international relations (27%), and public administration and management (16%). Lastly, QCA is concentrated in four areas – political science (34%), public administration and management (23%), environmental sciences and studies (20%), and international relations (13%).

In a nutshell, the diversity of quantitative and qualitative methods across different research areas in social sciences confirms that there is no one best way to conduct a systematic CPA. The nature of the object of comparison and the challenge of coping with differences across political systems, policy areas, and historical contexts force researchers to consider different types of evidence, thereby giving up the single-method research design that might be sufficient in other fields. How can scholars then preserve the coherence of their research design when combining different methods? This question will be addressed in the following section.

4. RESEARCH DESIGN AND METHODS SELECTION

While CPA may well follow the mainstream regarding methodological developments of the social sciences, scholars involved in this field may find some advantage in utilizing a broader array of methods to deal with the complexity inherent to public policies, and the need to understand the roles played by process and context in case selection. This requires solving the classical problem of alignment between ontology and methodology, which will guide scholars' election for an extensive or an intensive research design.

4.1 Aligning Ontology and Methodology

The definition of a strategy to answer a research question is a matter of methods and techniques, but the definition of a criterion to choose a method is a matter of methodology (Sartori 1970). A methodology is best understood as the product of a philosophical ontology (the relationship between the mind and the world) and a scientific ontology (the relationship between the status of empirical data and our knowledge about the world) (Jackson 2016).

This “stratified ontology” (Sayer 2000: 11–12), grounded in the distinction between the real, the actual, and the empirical, stems from the difference between the intransitive and the transitive dimensions of knowledge, that is to say between the physical processes and social phenomena in the world, on one hand, and the discourses and theories about the world, on the other hand. On one hand scholars' philosophical ontology can be dualist, if they conceive

Table 1.1 *Four methodologies*

		Scientific ontology	
		Phenomenalist	Transfactual
Philosophical ontology	Dualist	Neo-positivism	Critical realism
	Monist	Analyticism	Reflexivity

Source: Jackson (2016).

a gap between the mind and the world, or monist, if they conceive the world as a continuity of the mind. On the other hand, their scientific ontology can be phenomenalist, if their knowledge about the world is based exclusively on empirically observable data, or transfactual, if their knowledge about the world is also based on detectable but non-empirically observable data.

The stratified ontology supports a four-category typology of methodologies, that goes way beyond the somehow sterile opposition of quantitative versus qualitative methods (Table 1.1). Type 1 (neo-positivism) combines dualism with phenomenism, type 2 (realism³) combines dualism with transfactualism, type 3 (analyticism) combines monism with phenomenism, and type 4 (reflexivity) combines monism with transfactualism (Jackson 2016). On one hand, realists and neo-positivists share a dualist ontology, so they concur in that causal explanations of the world are possible. Conversely, analyticists and reflexivists share a monist ontology, so they concur in that only interpretations of the world can be provided. On the other hand, neo-positivists and analyticists share a phenomenalist ontology, so they concur that the scientific knowledge about the world is exclusively based on directly observable data. Conversely, realists and reflexivists share a transfactual ontology so they concur in that their knowledge is also based on detectable but non-directly observable data.

On one hand, types 1 and 2 share a dualist ontology, so they concur in that causal explanations of the world are possible. Conversely, type 3 and 4 share a monist ontology, so they concur in that only interpretations of the world can be provided. On the other hand, types 1 and 3 share a phenomenalist ontology, which means they concur that the scientific knowledge about the world is exclusively based on directly observable data. Conversely, types 2 and 4 share a transfactual ontology, which means they agree in developing scientific knowledge with detectable but non-directly observable data.

Aligning these methodologies with our methods in CPA requires a clear understanding of the logic of causation they support, as well as their implications for the aims of intervention, the modalities of evaluation, and the outputs for policy analysis and design (Table 1.2). According to this typology, type 1 research advocates for a logic of causation based on variable-oriented explanation, aimed at falsifying law-like theories and predictive models through the counterfactual analysis of observable data. Type 2 research favors case-oriented explanations to build and test middle-range theories and contingent patterns grounded in the transfactual analysis of non-observable data. Type 3 research would rather adopt a logic of causation based on a variable-oriented interpretation to build and test ideal-type theories and typologies through the counterfactual analysis of observable data. Type 4 research values case-oriented interpretation for critical theory-testing and narratives based on the transfactual analysis of non-observable data.

Classical examples of theories grounded in these methodologies are the punctuated equilibrium theory – a law-like theory – for type 1 (True, Jones, and Baumgartner 2007), the logic of appropriateness – a middle-range theory – for type 2 (March and Olsen 2006), the social construction of policy design – an ideal-type theory – for type 3 (Schneider and Ingram 1997), and

Table 1.2 Methodological implications of different ontologies for CPA

Methodology	Logic of causation	Aims of intervention	Modalities of evaluation	Outputs for policy design
Type 1: Neo-positivism	Variable-oriented explanation	Law-like theories falsification	Counterfactual analysis of observable data	Predictive models
Type 2: Realism	Case-oriented explanation	Middle-range theory-building and testing	Transfactual analysis of non-observable data	Contingent patterns
Type 3: Analyticism	Variable-oriented interpretation	Ideal-type theory-building and testing	Counterfactual analysis of observable data	Typologies
Type 4: Reflexivism	Case-oriented interpretation	Critical theory-testing	Transfactual analysis of non-observable data	Narratives

Source: Adapted from Jackson (2016) in: Fontaine, Medrano, and Narváez (2019).

the argumentative turn in deliberative policy analysis – a critical theory – for type 4 (Fischer 2003). Of all the reasons that have made these theories robust explanations or interpretations of policy processes, the most noteworthy here is that their authors proceeded with a careful alignment between ontology and methodology. This means that each one of these theories is bounded to a particular combination of philosophical and scientific ontology, which commands a particular conception of causation and causality, which in turn commands a particular combination of methods and techniques.

Therefore the four types of methodologies should be used as guidelines to secure the external coherence and the internal consistency of a research design, when it comes to the choice of a method or the combination of multi-methods. This leaves us with two different research designs – extensive or intensive – as explained below.

4.2 Extensive versus Intensive Research Design

Since they vary across the methodologies described above, the meanings of causality (a causal relationship between two entities or events) and causation (a hypothesis or a theory about a causal relationship) raise three intertwined problems: regularity, necessity, and symmetry. Regularity refers to a causation based on multiple observations, which is essentially a question of quantitative measurements. Hence it lies at the core of probabilistic methods and is generally associated with symmetry and correlation (the Holy Grail of statistics), which means that a positive or negative variation of a cause is correlated with the symmetric variation of its effect. Necessity refers to a causation based on a detailed account of the factors involved in a relationship, which is essentially a qualitative problem. It is at the heart of deterministic methods, which are generally concerned with the positive dimension of a causal relationship – the actual effect of a causal force – rather than with symmetry.

The major difference here is that establishing regularity in a causal relationship is about the mean causal effect based on as many observations as possible, while establishing necessity in a causal relationship is about the evidence that connects a cause to an effect based on a few cases or even a single-case study. Confirming or disconfirming regularity requires an extensive research design, while confirming or disconfirming necessity commands an intensive one (Sayer 2000). Both types of causation are actually complementary. For instance, based on a large-N

data panel we could observe a correlation between the event A (smoking) and the event B (dying of lung cancer), but this actually tells us nothing about how this particular cause produces this particular effect. Therefore a full-length theory of A causing B would require a multi-method research design combining an extensive approach (to establish the mean causal effect caused by A on B) with an intensive one (to establish the causal mechanism linking A to B).

Still, each part of such a research design aims at confirming or disconfirming different kinds of theories. An extensive research design may produce a causal law or a general theory. It aims at dealing with causality as a formal relation of similarity, which leads to asking which conditions are theoretically necessary to trigger a process. In open systems – such as public policies – this can only be done with semi-experiments (by manipulating the real world) or with statistics (in search of mean causal effects). Conversely, an intensive research may produce an instrumentalist law or a middle-range theory. It seeks to address causality as a substantial relation of connection, which leads to asking what it is about an entity that produces a causal power. This can only be done through deep within-case study or small-N comparison (in search of causal mechanisms).

The difference between intensive and extensive research designs does not overlap the dilemma between quantitative and qualitative methods (as a matter of fact they may both combine these methods). Yet intensive research is akin to case study analysis and small-N comparison, unlike extensive research which requires large-N comparison occasionally completed by case studies. In either case, multi-methods are utilized for exploratory purposes in theory-building and for confirming or disconfirming purposes in theory-testing. But only intensive research may be interested in deep within-case studies of non-representative processes.

Eventually, this division of labor refers to different approaches to the issue of theoretical generalization and different appreciations regarding the number of cases or observations deemed sufficient to build and test a causal explanation or interpretation. It is obviously risky for the coherence and the consistency of a research design to combine both intensive and extensive approaches in a single research design, without having previously determined what kind of causation they are expected to confirm or disconfirm. Therefore in practice, regularity is often grounded in the phenomenalism of types 1 and 3 methodologies, while necessity is more commonly grounded in the transfactualism of types 2 and 4.

Moreover, for types 1 and 3, the external validity of a causal relationship depends on the correlation between an independent and a dependent variable, so that the higher the number of observations, the better for the research. From that standpoint, most scholars consider qualitative methods inaccurate to establish or test a causal relationship so they utilize them, at best, as a complement to statistics. For instance they may use process tracing as a secondary technique, to shed light on a particular aspect of probabilistic models built on econometrics (Collier 2011; Seawright 2016). Even when they value the probatory contribution of case studies (Goertz and Mahoney 2012; Goertz and Starr 2003; Mahoney 2001) their search for sufficient and necessary conditions still requires the kind of cross-case comparison driven by a probabilistic logic of causation. For instance they may combine process tracing with comparative historical analysis and political economy (Falletti 2010; Hall 2012; Mahoney 2012).

4.3 The Problem of Case Selection

This handbook focuses more on small-N research design than most discussions on comparative methods in social sciences. For the reasons discussed already, such as the importance of context, we consider it crucial to think about how to conduct a research that looks carefully at a limited number of cases. The obvious example of the kind is the case study, but other methods such as ethnography also tend to focus only on a few cases, as they attempt to understand them more completely than might be possible through a large-N research design.

Once we decide to conduct a small-N comparison, a number of subsidiary questions arise about how this is to be done. The first is, which case or cases are to be considered? This question is addressed in some greater detail in the chapters on the comparative method and the most-similar and most-different systems designs, but here we point out that cases are only to some extent natural occurrences – cases have to be *constructed* by the researcher (Ragin and Becker 1992). Further, scholars and practitioners must decide whether their case is a case of some particular process or attribute. The same set of facts can be used in a variety of different ways to address different research questions.

A second dilemma raised by case selection is whether to use most-similar or most-different cases for comparison. The usual approach among social scientists has been to pick most-similar cases, and to use their similarity as a means of controlling extraneous variance. Then again, it is up to the researcher to make the case for similarity, and cases that are most-similar for some purposes could be most-different for others. Some scholars have argued in favor of utilizing most-different case designs (Przeworski and Teune 1970), with the argument being that if a relationship holds up across most-different cases then it is robust and can be seen almost as a law of political behavior.

Another issue which arises in case selection is the possibility of identifying a crucial case (Eckstein 1975), or a case that can be used to make a definitive test of a proposition. The design is, for example, to find a case in which the hypothesized relationship is least likely and determine if it is, in fact, supported. Some have argued that the possibilities of identifying these crucial cases are remote and consequently we should abandon that pursuit in favor of “pathway cases” (Gerring 2007) that can be used to elucidate the causal mechanisms involved in a theory rather than confirming or disconfirming that theory.

Finally, there is a question of linking within-case and cross-case analysis. Most case study methods focus on examining relationships and testing propositions within the single case. But for comparative analysis it is important to compare patterns across cases. Cross-case CPA allows us to understand the effects of context on a causal relationship, and to take into account other possible explanations for the findings in any one case. This cross-case comparison is central to Lijphart’s comparative method, which he identified as one of the four fundamental research methods in the social sciences.

In a nutshell, each research design raises a specific question, which requires a specific set of methods to be aligned with a type of methodology according to the meaning of causality and causation. A general theory of causation describes regularities but it does not identify causal forces so it is not explanatory in a deterministic sense. A middle-range theory of causation describes a necessity, hence it explains how the qualitative nature of a social phenomenon varies according to the context, but its generalization is limited by contingency. Understanding the complementarities of these theories is paramount, if CPA scholars wish to preserve the coherence and consistency of their research design, especially when opting for

a multi-methods strategy. This applies both to dualists, who seek to produce explanations, and monists, who contend that only interpretations are in the realm of the social sciences. It also applies both to phenomenologists, who consider the only actually scientific knowledge is based on directly observable data, and transrationalists, who contend that this knowledge is also based on detectable though non-directly observable data.

5. CONCLUSION: THE IMPORTANCE OF METHOD FOR COMPARATIVE POLICY RESEARCH AND PRACTICE

We shall conclude this introductory study by returning to some of the central contentions of the handbook. To characterize the current diversity of methods at the disposal of CPA scholars, it appears that the same biases prevail that exist in comparative politics, and in social sciences generally speaking. There is a persistent orthodoxy, favoring quantitative analysis and rejecting the possible contributions of qualitative techniques and multi-methods. Many high-impact journals these days are unfortunately filled with research articles that have followed a neo-positivist orthodoxy, hence failing to consider the full range of options and, perhaps more important, often failing to think about triangulation and the use of multi-methods as a means of gaining a more complete picture of the policy issue under scrutiny.

Yet at the same time it has become entrenched, there are more heretics providing alternatives to the received wisdom, whose heresies are supported by the existence of powerful qualitative methods, but perhaps even more by the increased recognition of the need to understand more about the process of policy making and the role of somewhat amorphous concepts – such as culture – on policy. Considering its multi-disciplinary nature, CPA is arguably more akin to multi-methods than any other social sciences areas. For one thing, context is important and bringing it in may effectively involve more qualitative understanding (Pollitt 2013). Moreover, CPA often involves a detailed examination of the policy process, which almost inherently requires qualitative methods such as congruence analysis and process tracing, small-N comparison, ethnography, etc.

As the content of this handbook indicates, we are ourselves very open-minded about which methods to use in CPA and above all we intend to make a modest but decisive contribution to the methodological debate. The book is organized in six parts. **Part I** is dedicated to the current methodological debate in social sciences that interests CPA. **Part II** draws on the contemporary trends in the methods and research agendas of some major analytical frameworks, related to comparison and causality in policy studies. **Part III** raises the problem of measurements and experiments in CPA. **Part IV** deals with mixed- and multi-methods in CPA research designs. **Part V** presents outstanding qualitative techniques more and more utilized in CPA. **Part VI** offers further reflections on the development of the field and the methodological debate thereof.

What we are advocating, above all, is a careful and conscious choice of the methods to be used, given the strength that triangulation of methods, and theory, can bring to research. On their own, all methods are not inherently good or bad, powerful or weak. What matters is under what conditions they are used and how well they are applied. The conscious choice of methods is important inasmuch as these have a decisive influence on our findings. Methods are not neutral instruments that will produce “correct” results, but they rather have their own biases and their own blind spots. For example, most quantitative methods assume an

additive relationship among variables, while QCA assumes a more interactive relationship, in which combinations of variables are associated with outcomes, and the absence of one or another produces no result, rather than merely a weaker result. Likewise, within-case study techniques look for particular decisions that produce the final outcome, rather than assuming that an outcome is a conjunction of variables. Both QCA and within-case methods are more deterministic than stochastic, hence they will provide a different type of understanding of the relationships among variables than will additive, statistical methods.

The methodological issues become more pressing when we attempt to explain differences across units, be they states, provinces, or whatever. As is discussed in several chapters of this handbook, case selection is crucial for conducting good comparative research, whether on public policy or any other topic. Cases picked simply out of habit or convenience are unlikely to provide the type of control over variance that is crucial for comparative analysis. Likewise, measurement is made more difficult by engaging in comparative analysis. Concepts may not “travel” well in comparative research (Sartori 1970), and that is even more of an intense problem for the measurement of those concepts, whether the measurement is qualitative or quantitative. A useful indicator in one context may be meaningless in another social system or culture, so the comparative researcher must pay particular attention to the meaning and context of the instruments used to measure.

The catalog of methods and techniques presented here should make it clear that there is a rich array of possible choices for the researcher. This handbook provides a good description and assessment of the methods available for CPA at this time. There are other methods that could have been included, but we do cover those most commonly in use for comparative analysis. Ultimately, this field is changing fast (see **Chapter 21** by Geva-May, Hoffman, and Muhleisen), so we need to be aware that those changes will affect the choice of methods. As already noted, the expanding use of experimental methods and the general development of qualitative techniques are enriching the methodological debate on CPA. The future for research appears extremely bright given the continuing development of techniques for understanding policy and the policy process. But, since choice can also be difficult, scholars must understand a range of alternatives in order to make the most appropriate choices for the problem at hand.

NOTES

1. Source: WOS database accessed on April 11, 2019.
2. N=79,471. Source: WOS database accessed on April 11, 2019.
3. The “critical” realist argument against Hume’s “empirical realism” is that the world is an open system where regularities are more the exception than the rule, and our knowledge about the world cannot be reduced to atomistic events (Archer et al. 1998; Sayer 1992). Nevertheless we refer to “realism” instead of “critical realism”, thus following the original statement of a “realist scientific endeavor” (Bhaskar 1978). The opposition of a “critical” realism to Hume’s “empirical” realism is theoretically misleading, as it should apply to all but one methodologies. As a matter of fact scholars grounded in a reflexivist methodology owe a great deal to the critical theory which gave initially way to critical realism.

REFERENCES

- Archer, M., Bhaskar, R., Collier, A., Lawson, T., and Norrie, A. (Eds.) (1998). *Critical Realism: Essential Readings*. London: Routledge.
- Bardsley, N. (2010). *Experimental Economics: Rethinking the Rules*. Princeton, NJ: Princeton University Press.
- Beach, D. and Pedersen, R. B. (2016). *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching and Tracing*. Ann Arbor, MI: University of Michigan Press.
- Berg-Schlosser, D. (2013). *Process Tracing Methods: Foundations and Guidelines*. Ann Arbor, MI: University of Michigan Press.
- Bhaskar, R. (1978). *A Realist Theory of Science*. London: Routledge.
- Blatter, J. and Haverland, M. (2012). *Designing Case Studies: Explanatory Approaches in Small-N Research*. Basingstoke: Palgrave Macmillan.
- Blatter, J. and Haverland, M. (2014). Case studies and (causal-) process tracing. In I. Engeli and C. Rothmayr Allison (Eds.), *Comparative Policy Studies: Conceptual and Methodological Challenges*. Basingstoke: Palgrave Macmillan, 59–84.
- Brady, H. E. and Collier, D. (2010). *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield.
- Cairney, P. (2016). *The Politics of Evidence-Based Policy Making*. London: Palgrave Macmillan.
- Carter, P. (2012). Policy as palimpsest. *Policy & Politics* 40: 423–43.
- Castles, F. G. (2004). *The Future of the Welfare State: Crisis Myths and Crisis Realities*. Oxford: Oxford University Press.
- Collier, D. (2011). Understanding process tracing. *Political Science and Politics* 44 (4): 823–30.
- Dunlop, C. A. and Radaelli, C. M. (2013). Systematizing policy learning: from monolith to dimensions. *Political Studies* 61: 599–619.
- Eckstein, H. (1975). Case studies and theory in political science. In F. I. Greenstein and N. W. Polsby (Eds.), *Handbook of Political Science*, vol. 7: *Political Science: Scope and Theory*. Reading, MA: Addison-Wesley, 79–133.
- Engeli, I. and Rothmayr Allison, C. (Eds.) (2014). *Comparative Policy Studies: Conceptual and Methodological Challenges*. Basingstoke: Palgrave Macmillan.
- Falleti, T. (2010). Infiltrating the state: the evolution of health care reforms in Brazil 1964–1988. In J. Mahoney and K. Thelen (Eds.), *Explaining Institutional Change: Ambiguity, Agency and Power*. Cambridge and New York: Cambridge University Press, 38–62.
- Fischer, F. (2003). *Reframing Public Policy: Discursive Politics and Deliberative Practices*. New York: Oxford University Press.
- Fontaine, G., Medrano, C., and Narváez, I. (2019). *The Politics of Public Accountability: Policy Design in Oil-Exporting Countries from Latin America*. Cham: Springer.
- Gerring, J. (2007). Is there a (viable) crucial case method? *Comparative Political Studies* 40: 231–53.
- Gerring, J. (2010). Mere description. *British Journal of Political Science* 42: 721–46.
- Goertz, G. and Mahoney, J. (2012). *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.
- Goertz, G. and Starr, H. (2003). Introduction: necessary condition logics, research design, and theory. In G. Goertz and H. Starr (Eds.), *Necessary Conditions: Theory, Methodology, and Applications*. Lanham, MD: Rowman & Littlefield, 1–24.
- Hall, P. (2012). Tracing the progress of process tracing. *European Political Science* 12: 20–30.
- Hogwood, B. W. and Peters, B. G. (1987). *Policy Dynamics*. Brighton: Wheatsheaf.
- Jackson, P. T. (2016). *The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics* (2nd edition). London: Routledge.
- Kellstedt, P. and Whitten, G. (2013). *The Fundamentals of Political Science Research*. New York: Cambridge University Press.
- King, G., Keohane, R. O., and Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Mahoney, J. (2001). Path-dependent explanations of regime change: Central America in comparative perspective. *Studies in Comparative International Development* 36 (1): 111–41.

- Mahoney, J. (2012). The logic of process tracing tests in the social sciences. *Sociological Methods & Research* 41 (4): 570–97.
- March, J. and Olsen, J. (2006). The logic of appropriateness. In M. Moran, M. Rein, and R. E. Goodin (Eds.), *The Oxford Handbook of Public Policy*. New York: Oxford University Press, 689–708.
- Pawson, R. (2006). *Evidence-Based Policy: A Realist Perspective*. London: Sage.
- Peters, B. G. (2013). *Strategies for Comparative Research in Political Science*. Basingstoke: Palgrave Macmillan.
- Peters, B. G. and Zittoun, P. (2016). *Contemporary Approaches to Public Policy*. London: Palgrave Macmillan.
- Pollitt, C. (2013). *Context in Public Policy and Management: The Missing Link?* Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Przeworski, A. and Teune, H. (1970). *The Logic of Comparative Social Inquiry*. New York: John Wiley.
- Ragin, C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of Chicago Press.
- Ragin, C. C. and Becker, H. S. (1992). *What is a Case? Exploring the Foundations of Social Inquiry*. Cambridge: Cambridge University Press.
- Rihoux, B. and Grimm, H. (Eds.) (2006). *Innovative Comparative Methods for Policy Analysis*. New York: Springer.
- Rose, R. (1993). *Lesson Drawing in Public Policy: A Guide to Learning Across Time and Space*. Chatham, NJ: Chatham House.
- Sartori, G. (1970). Concept misformation in comparative research. *American Political Science Review* 64: 1033–53.
- Sayer, A. (1992). *Method in Social Science: A Realist Approach* (2nd edition). London: Routledge.
- Sayer, A. (2000). *Realism and Social Science*. London: Sage.
- Scharpf, F. W. (1997). *Games Real Actors Play: Actor-Centered Institutionalism in Policy Research*. Boulder, CO: Westview Press.
- Schneider, A. and Ingram, H. (1997). *Policy Design for Democracy*. Lawrence, KS: University Press of Kansas.
- Seawright, J. (2016). *Multi-Method Social Science: Combining Quantitative and Qualitative Tools*. Cambridge: Cambridge University Press.
- Shedish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth.
- Starke, P. (2006). The politics of welfare state retrenchment: a literature review. *Social Policy & Administration* 40: 104–20.
- True, J. L., Jones, D., and Baumgartner, F. R. (2007). Punctuated-equilibrium theory: explaining stability and change in public policymaking. In P. A. Sabatier (Ed.), *Theories of the Policy Process*. Boulder, CO: Westview Press, 155–88.
- Van Kersbergen, K. and Vis, B. (2014). *Comparative Welfare State Policies*. Cambridge: Cambridge University Press.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. (1999). *Unobtrusive Measures*. Newbury Park, CA: Sage.
- Weible, C. M. and Sabatier, P. A. (2017). *Theories of the Policy Process* (4th edition). Boulder, CO: Westview Press.
- WOS (Web of Science) (2019). ISI Database. Accessed April 2019 at: <https://clarivate.com>.

APPENDIX

*Table 1A.1 Methods used for CPA by substantive areas (1945–2018)**

Area	Total (N=79,471)	Quantitative				Comparative				
		Group 1	Group 2	Group 3	Group 4	Group 1	Group 2	Group 3	Group 4	
Environmental sciences and studies	0.19	0.28	0.18	0.11	0.17	0.13	0.16	0.13	0.07	0.19
Economics	0.12	0.08	0.38		0.17				0.07	
Public environment and administration	0.07	0.05	0.05	0.07	0.22	0.17	0.17	0.16	0.14	0.16
Political science	0.07	0.06	0.09	0.04	0.10	0.10	0.09	0.11	0.41	0.32
Education	0.04	0.06		0.11	0.09	0.11	0.16		0.05	0.05
Management	0.04	0.05	0.09						0.05	0.05
Geography	0.03	0.06		0.09	0.08		0.06			
Computer science	0.03									
International relations	0.01		0.03	0.08	0.06	0.05	0.06	0.05	0.23	0.12
Health policy	0.01					0.08		0.06	0.05	
Health care sciences	0.01					0.10		0.07	0.05	
Sociology	0.01			0.11	0.06					
Anthropology	0.01			0.18						
Linguistics	0.01			0.04						
Communication	0.01									
History	0.01				0.05					
Other	0.33	0.35	0.18	0.24	0.36	0.21	0.14	0.42	0.00	0.00
Total	0.67	0.64	0.82	0.76	0.63	0.79	0.85	0.58	1.00	1.00

Note: * All key words associated with the descriptor “policy” in the query among “articles topics”.

Source: Web of Science Database accessed April 22, 2019.

PART I

THE METHODOLOGICAL DEBATE

2. The comparative method and comparative policy analysis

B. Guy Peters

1. INTRODUCTION

This chapter¹ will explore the utility of the comparative method for comparative policy analysis, and will be arguing that this method can be especially useful for dealing with comparative public policy. Although learning about policy in any of our individual countries can provide some understanding of the dynamics of policy, on that basis it is difficult to develop general theoretical and analytic models of public policy. The world, on the other hand, provides a natural laboratory for the study of policy (and indeed any other social or political phenomenon) that enables researchers to build theory and to understand the conditions under which certain factors can influence outcomes with a greater chance for valid generalizations.

The contemporary approach to analyzing policy choices made by individual nations is that of attempting to find natural experiments (Dunning 2012), otherwise styled as quasi-experiments, and to analyze them as if there they were indeed experimental situations. That model for research, however, tends to be applied when a single country changes some attribute, e.g. New Zealand adopting proportional representation, and considers some dependent variable, e.g. number of parties in parliament, over time. Comparative analysis, on the other hand, is able to generate variance from multiple (albeit a small number of) cases which may have made different choices, or the same choice under different political and social conditions.

The comparative study of public policy is also crucial for the more practical aspects of policy analysis. One of the contemporary mantras about policy is “evidence-based policy making”, with a good deal of the evidence being used coming from other political and policy systems (Cairney 2016). If, however, the evidence that is being used for making policies is faulty, or those attempting to use it do not understand the context within which the findings were developed, then any implicit advice that may be derived from the findings has an increased probability of being faulty. Although comparative analysis is sometimes thought to be an arcane academic exercise, it can provide major practical benefits for policy makers.

While the general idea of comparative policy analysis is appealing, that type of analysis is also very difficult to do well. Too often the countries (or provinces, or cities, or whatever) that are selected are chosen on the basis of familiarity or proximity rather than for analytic reasons. On the other hand, comparisons that are undertaken on more analytic grounds may not utilize understanding of the nature of the political systems involved, or of more subtle differences among the cases that in the end will have a significant influence on the policy choices being made. Navigating between this Scylla and Charybdis for comparison is a difficult task for any researcher.

To this point I have been discussing comparative policy analysis in rather general terms, meaning basically any study that examines policies in two or more settings, and/or over time. Much of the comparative policy analysis has been, and continues to be done, using large-N

studies and statistical methods, attempting to find the correlates of policy choices, or perhaps the correlates of success or failure of policies. While those studies do provide a broad brush picture of the influences on policy choice, for public policy perhaps more than other areas of social research it is important to understand the internal dynamics of the process as well as the rather generalized picture derived from the correlational analysis. For policy studies it may be more useful to get into the specific cases in some detail rather than just examining relationships among often simply defined variables.

This chapter will focus on the comparative method. While one would hope that most research in the social sciences would be implicitly if not explicitly comparative, there is also an approach to analysis that is commonly referred to as the *comparative method*. This method depends on careful selection of cases as the means of explaining observed differences in the dependent variable. As implied above it is assumed (in experimental terms) that at least one of the cases selected in the analysis will have received the treatment while others will not. If there are then differences in the outcome then the argument (based on Mill's principle of difference, see Morlino 2018) is that the treatment is the presumptive cause of the difference on the dependent variable.

But the comparative method involves a clear need to utilize other methods and techniques to make its comparisons. The comparative method does not specify clearly how evidence is to be collected, only that the important element of the method is the selection of cases. Much of the use of the comparative method has involved doing parallel case studies (De Winter, Della Porta, and Deschouwer 1996), but one could imagine using matched time-series statistical analyses in carefully selected contexts to then use for comparative analysis. Experimental methods tend to discount the importance of context, but if a researcher were clever enough to assume that context does matter then even the same experiments done in different settings could be useful.

Therefore, this chapter will also point to how the comparative method as described by Lijphart leads on naturally to using other small-N techniques for research. Indeed, one could argue that the comparative method is to a great extent merely a preliminary sorting method for doing other types of research. The comparative method helps the researcher identify appropriate cases for analysis. The researcher must then utilize another method, e.g. process tracing (see Beach, Chapter 14 this volume) or perhaps ethnographic methods (Pacheco-Vega, Chapter 18 this volume) to investigate the cases selected.

2. MODES OF ANALYSIS IN COMPARATIVE POLICY STUDIES

In his seminal article on comparative analysis in political science Arend Lijphart (1971) argued that there are four basic research designs and methodologies for the social sciences – experimentation, statistical analysis, case studies, and the comparative method. All of these basic research methods have been used to provide insights into public policy in a variety of settings. As noted above, the statistical method is the most commonly used in the social sciences (see Scruggs and Allan 2006), although experimentation is becoming more common and is increasingly argued to be the “gold standard” for research design. And for policy studies somewhat more than for other parts of the social science there continue to be a large number of case studies (Blatter and Haverland 2014). There is an almost endless stream of studies

discussing in great detail policy x in countries y and z, all of which contribute to our collective knowledge but which do not often provide themselves the rigorous comparative analysis that would be needed for building theory.

For all of these four methods of analysis there is a fundamental litany for research design that should be addressed in the design (see Peters 2013). This litany is:

Maximize experimental variance, minimize error variance, and control for extraneous variance.

This statement is based on the measurement of the dependent variable. Any observation will have three components. The first is the “true” relationship between the (presumed) independent variable and the dependent variable – the experimental variance. There is also some possibility of error entering into the observation. This element may be especially high in qualitative research that depends heavily on the observations of an individual researcher, or a limited number of researchers, but also can be present in quantitative research.²

The third element of this litany is the most important for thinking about comparative research designs. Although we may observe a strong relationship between some variable x and some variable y, how do we know that this occurs only because they are both related to some z? That extraneous variance can confound any findings we have, whether developed through quantitative or qualitative methods. Coping with extraneous variance and spurious relationships is relatively easy for quantitative work – all one need do is to introduce control variables into the model and determine if the relationship between the independent and dependent relationship maintains its strength.³

In the experimental method the problems of extraneous variance are assumed to be handled through random assignment to test and control groups.⁴ In the case method there is little real protection against extraneous variance, other than the researcher being open to alternative explanations of his or her findings, and as much peer advice as possible. It is the comparative method, the focus of this chapter, where some of the more interesting questions about coping with extraneous variance arise.

For the comparative method, the selection of cases is the means through which extraneous variance is controlled, and indeed also the means through which the experimental variance is built into the analysis. The logic of the comparative method is that if the researcher selects the appropriate “sample” of cases – be they countries or sub-national units or time periods – then he or she can test hypotheses by ascertaining the way in which the relationship among variables operates within the several cases (see Seawright and Gerring 2008). The cases, of course, are not a sample in the usual sense of the term but a more purposive selection designed to control for extraneous conditions.

The use of selected cases to test hypotheses is important for theory-testing because it enables the researcher to bring in the role of context (Falleti and Lynch 2009). The classic work by Przeworski and Teune (1970) (see below) argued that one of the purposes of good comparative analysis is to eliminate the names of countries and replace those with variables. While I understand the scientific logic behind the statement, I would argue that this to some extent defeats the purpose of comparative analysis. We are interested in the names of countries not just for geographical reasons but because they reflect the context within which the processes in which we are interested function. These countries are bundles of variables that constitute the context and hence are crucial for understanding (see Pollitt 2013).

The selection of cases for the comparative method poses something of a dilemma for the researcher. On the one hand, the logic of the research design is to control extraneous variance. If so, then the cases selected should be as alike as possible on the factors that may be assumed to produce variations in outcomes. Using that logic the purpose of choosing cases carefully is, insofar as possible, to *eliminate* context. If this is done perfectly (almost impossible, of course) then context is virtually gone and we are left with the bundle of variables that Przeworski and Teune want for the analysis.

It is also important to note that similarity and difference among cases is often constructed, and the same two cases may be well-controlled in some circumstances and not in others. For example, the United States and the United Kingdom have been used together in countless studies assuming this controls variance, but for many, if not most, institutional variables they are quite different.

That role of context and contingency in comparative analysis moves away from the dominant logic of variable-based research to more case-based research. There are still variables operating within those cases, of course, but the context of the case tends to affect the manner in which the variables interact. Thus, to some extent, as well as controlling for extraneous variance the comparative method tends to build in that extraneous variance as a major component of the analysis. But that, in turn, presents significant difficulties in untangling the relevant and the less relevant elements of context.

When using the comparative method, the principal challenge for researchers is to select the cases for inclusion in the “sample” of countries. The comparative method is a small-N approach that seeks to use case selection to control a number of variables in the cases selected. This control is the analog of inserting control variables into a regression equation, and is intended to isolate the effects of a presumed independent variable, or perhaps several independent variables, on the dependent variable – usually a policy choice (see van den Heijden 2014). Unlike other methods, the selection of cases is not random, but rather is designed to demonstrate something about policies in the several cases.

Having said the above, the obvious question becomes which cases to select? And that question involves several subsidiary questions, each of which is important for the analysis and the validity of the findings. The first question is a general research design question about the nature of the “sample” to be analyzed. Most comparative policy papers tend to adopt a “most-similar systems” design, selecting cases that are as similar as possible, but yet vary on some key feature or features (see Anckar, Chapter 3 this volume). This selection may not be done explicitly by the individual doing the research, but may simply be the product of the range of knowledge or language skills of the researcher. The logic here is that the selection of similar cases is a means of controlling for extraneous variance that may confound the relationship among variables.

The logic of “most-different systems” design is very different. By choosing cases (whether national or sub-national) that are extremely different on a range of variables, the assumption is that if a relationship is found between some independent and some dependent variable then it is more likely to be “true” than if the same relationship were found in a more constrained set of cases. The use of most-different systems designs, however, tends to depend upon having very clearly defined hypotheses about behavior, and generally works better with large-N studies. Indeed, the original logic of most-different systems was for it to be used on individual-level data, rather than the macro- or meso-level data more common for comparative policy studies.

Table 2.1 *Purposes of comparative research*

Purpose of comparison	Treatment of societies	
	Society as an entity	Society as a set of conditions
To show similarities	To identify universals	To prove general propositions
To show differences	To describe why societies are different	To specify time–space coordinates of propositions

Source: Allardt (1990).

Hence, the most-different systems strategy tends to be less useful for comparative policy analysis than for other aspects of the social sciences (Yom 2015).

The most-different systems design also poses problems of conceptualization and measurement to a much greater extent than does the most-similar design. Giovanni Sartori's identification (1970; see also Gerring 1999) of the "traveling problem" in comparative research remains important, and is especially relevant for most-different designs. Concepts that are familiar and meaningful in some settings may not be in others, especially when research is conducted in different cultural and political environments. Even policy terms that appear to travel across countries may have different connotations in different settings.

To this point I have been discussing comparative analysis as analysis based on geographical units. This is certainly the most common form of comparative analysis, but we should also think about comparisons across policy areas. As Gary Freeman (1985) argued, the differences across policy areas may be more significant than the differences across countries, and certainly may be more different than differences across sub-national units within a particular country. By examining policy differences researchers can ascertain to what extent the policy process within a geographical unit is consistent across policy areas, as well as understanding the consequences of particular policy problems for that process.⁵

The sociologist Erik Allardt provides an interesting analysis of how to think about the comparative research process, based in part on the Przeworski and Teune differentiation between most-similar and most-different designs. Allardt raises the fundamental point (see Table 2.1) of what the intention of the comparative research might be. Much of the discussion coming from Przeworski and Teune, and many other scholars in comparative politics and sociology, is that the purpose of the research may be to test theory and to "prove" generalizations. As shown in the table this is but one of four possible uses of comparative analysis, and for comparative policy studies may not even be the most important.

For comparative policy studies two of the other cells in this table may be of particular relevance. One of these is the identification of universals. If, as noted above, one of the important uses of comparative policy research is to facilitate policy transfer and "evidence-based policy making" (Cairney 2016) then identifying the universals in policy systems is crucial. Without understanding the commonalities in policy making, and in the success and failure of policy, then that transnational diffusion is more likely to fail.

The second useful aspect of comparative analysis other than testing general propositions is to develop propositions that are more limited in time and space. As argued above, an understanding of context is crucial for comparative analysis, and without that understanding again any capacity to use the results of comparative analysis either for theory or for more practical concerns is limited. This need to understand context, and integrate context into the analysis, may be especially relevant for comparative policy studies, given that policy often has a signifi-

cant cultural and social element without which the researcher may well not understand the true import of the findings. While this may appear to be the antithesis of proving generalizations, being able to understand when and how propositions are *not universal* may actually add more to our knowledge of comparative politics and policy.

3. SAMPLING ON THE DEPENDENT VARIABLE

Another issue which arises in comparative research that is based on selecting cases purposefully is the possible selection on the dependent variable. This is perhaps a natural tendency. We want to know why a particular type of policy works, so we look at successful cases. But Barbara Geddes (2003) refers to this as an “inferential felony”, given that if there is no variance of the dependent variable we have no way of knowing what factors are associated with success or failure. That said, however, sampling on the dependent variable does permit us to eliminate some potential factors from consideration,⁶ and can provide some initial understandings of the political patterns associated with success or failure.

The capacity to eliminate possible necessary conditions through selection of the dependent variable may be especially important for comparative policy analysis. Given that success in policy may depend upon a large number of factors, some of which may be beyond the control of government, understanding what factors to emphasize in making an intervention becomes especially important. Likewise, for understanding policy choices and outcomes from an academic perspective, eliminating necessary conditions permits the researcher to make more precise specifications of the relationships.

It is important to note here that although I have been speaking in terms of variables, the same logic applies for qualitative as well as quantitative research. A variable is simply an attribute that varies, whether it is measured quantitatively or qualitatively. The same basic rules of research design apply in all these cases (see Peters 2013). While the qualitative measures may lack the *apparent* precision of the quantitative research, if done properly the qualitative measures can be as valid as the quantitative. The same standard for research – that the means of performing the measures be reproducible and intersubjectively transmissible – applies to both forms of measurement.

While quantitative methods may appear to provide more powerful means to understand comparative policy phenomena, qualitative methods provide powerful, if very different, mechanisms for explanation. While statistical methods are probabilistic and additive, qualitative methods tend to be more deterministic and to rely on the configurations of variables (Berg-Schlosser et al. 2009). Thus, qualitative methods tend to focus on the existence of necessary and sufficient conditions, and configurations of variables rather than the relative contributions of variables to the explanation. That naturally leads on to the utilization of methods such as qualitative comparative analysis (QCA) that help to identify the existence of such conditions.

We can argue that techniques such as QCA are built on some of the underlying assumptions of the comparative method. While the reliance on a small number of matched cases in the comparative methods precludes the search for configurations involving multiple variables and a medium sized N, the search for necessary and sufficient conditions for the occurrence of a phenomenon is similar. The comparative method may only use paired cases, or a very

limited number of cases, but both attempt to eliminate potentially spurious causes by matching the cases as closely as possible.

4. THE COMPARATIVE METHOD AND CASE STUDIES

From the above discussion it should be clear that doing good comparative research, whether on public policy or anything else, involves a great deal of knowledge even before the more formal parts of the research process begin. Selecting cases requires the researcher to know a good deal about them in advance, so that every strategy that is being followed can be effective. And even then surprises are likely to occur, given the complexity of the political, social, and economic circumstances of almost all countries. The comparative method in use may therefore require a good deal of trial and error, while the reconstructed logic may make the research appear well designed from the beginning.

This brings us to the point that having said that the comparative method depends upon case selection, what do we do after we have selected the cases? Much of the application of the comparative method involves doing case studies within cases that have been carefully selected as either most-different, or most-similar, cases, or perhaps simply because they are interesting. Especially for comparative policy analysis the comparative method may involve using methods such as process tracing to understand how policies are being made (Blatter and Haverland 2014; see also Blatter and Blume 2008).

Just saying that then the researcher does several comparable case studies to ascertain the relationship among variables is perhaps easier said than done. Making the comparison will involve detailed research in the two cases and with that a significant depth of understanding of those cases. And, as with any case studies there is the danger that the commitment of a researcher to a particular theory will determine, if unwittingly, the outcomes. Despite the richness of case studies, especially when several are done utilizing the same set of propositions, there are always important questions of internal validity.

5. THE POSSIBILITY OF HYBRID DESIGNS

I have been discussing the four methods proposed by Lijphart, and especially the comparative method, individually. Each of the methods does indeed have its utility and can reveal important aspects about public policy in a variety of settings. But we should also consider ways of combining these methods, and especially linking the more quantitative methods – experimentation and statistics – with the more qualitative methods – comparative and case studies. This is hardly a novel idea (see Abadie, Diamond, and Hainmueller 2015; Lieberman 2005) but it needs to be considered specifically for the analysis of public policies. Further, much of the discussion of moving iteratively between large-N and small-N analysis has dealt with those manners of research in rather general ways, rather than addressing specifically the place of the comparative method in these hybrid research designs.

For our purposes in this chapter perhaps the most promising form of interaction among research designs may be between experimental and comparative designs. Take, for example, the widespread introduction of programs of public management reforms since the 1980s. These interventions can be seen as experiments in which the “treatment” is initiating a reform

such as agencification (Verhoest et al. 2012). We know well that this reform has had varying effects across the numerous countries in which it has been tried, but why? There are numerous case studies of success and failure, as well as a limited number of more comparative analyses, but if these were conceptualized more as a set of natural experiments, or again as the quasi-experiments of Donald Campbell, the findings could be made more powerful. That is, they could be more powerful provided that the effects of context could be integrated effectively.

Further, as noted already, the case method and the comparative method are almost inherently related. One might think of the comparative method as doing several linked case studies and then comparing the results as the means of making the comparison. This linkage is especially strong when the case research is done using a theoretically informed process tracing approach, but is useful even when using less sophisticated versions of the case method.

One standard question which arises in these attempts to integrate quantitative and qualitative methods is whether the role of the qualitative analysis is to generate hypotheses or to test hypotheses. Given the dominance of quantitative analysis in the contemporary social sciences, the usual answer to that question is that while qualitative analysis is adequate for generating hypotheses, the heavy lifting of testing hypotheses should be left to quantitative analysis. Even with the development of stronger qualitative methodologies such as the numerous variants of QCA, the assumption is that only quantitative analysis can provide a true test of relationships among variables, especially one that is generalizable.

We could, however, make an alternative argument that while quantitative analysis provides identification of broad patterns of relationship among variables, qualitative analysis can provide more definitive testing of relationships. The quantitative analysis is, of course, stochastic and generates probabilities of a true relationship. Careful qualitative analysis, on the other hand, can be deterministic. Even if one does not accept the logic of the crucial case (Gerring 2007), attempting to find cases in which a particular hypothesis is less likely to be supported can be a powerful means of testing hypotheses and theory.

In Lieberman's often discussed model of nested analysis (2005) the assumption is that the researcher should begin with a quantitative analysis, determining the general strength of the relationship among the variables. The quantitative analysis will also permit identifying the deviant cases (the large residuals from the regression line). Doing case analysis of those deviant cases will then allow the researcher to understand why those cases were deviant and also give insights into the relationships among variables. Based upon that qualitative analysis the quantitative analysis can be refined and tested again.⁷

If we move away from the familiar quantitative versus qualitative divide in the social sciences, we must confront the relationship between comparative analysis and case studies. In most of the discussion above these approaches to research have been treated separately, but in the actual practice of research they can be, and perhaps should be, closely linked. It may be possible to link the comparative method with other modes of research such as experimentation (see Beach and Rohlfing 2018) but for most comparative scholars (and perhaps especially for those in comparative policy studies) the more natural linkage is with case studies.

6. SEVERAL THOUGHTS ON SMALL-N RESEARCH

The contemporary social sciences tend to focus on large-N quantitative studies, and increasingly on experimental designs, to attempt to identify causal relationships among variables. In the former there is a search for an average causal effect of a presumed independent (causal) variable across a range of dependent variables. In the latter the search is for treatments representing causal variables that can produce significant differences between treatment and control groups in a more or less controlled environment.

These emphases appear to leave small-N research, including research based on the comparative method, out of active consideration within the discipline (see Toshkov 2018). The argument that has been advanced is that small-N research is viable in demonstrating causality only when cross-case (following the application of the comparative method one assumes) is combined with within-case analysis. But this is precisely the argument being made above. The comparative method can tell the researcher which cases may be suitable for comparison if one is interested in determining the effects of a certain presumed causal variable, but there is still the need for the within-case analysis.

But one may be able to make an even stronger argument for small-N research, and especially for the within-case analysis. Mahoney (2007) and others have pointed out that when looking for causal mechanisms it may be better to engage in careful analysis of a limited number of cases rather than examine the plentitude of cases in large-N studies. The findings about causation can be tested with large-N analysis. This reverses the usual assumptions about how to do multi-method, or nested analysis (Lieberman 2005), and assigns a more significant role to the small-N component of research.

We also need to remember that determining causal relationships may not be the only purpose for doing comparative policy research. That is certainly the primary goal in the canon of the social sciences, but small-N research is also useful, for example, for identifying necessary conditions for the occurrence of some phenomenon. That determination may be very useful for designing public policies, and can be the precursor to research that focuses more directly on causation.

7. CONCLUSION: WHAT, IF ANYTHING, IS SPECIAL ABOUT COMPARATIVE POLICY STUDIES?

This chapter has discussed the comparative method and comparative research in somewhat general terms, albeit with some allusions to comparative policy research. I have been arguing that in general the comparative method is a powerful mechanism for organizing research on a range of social phenomena, given that it can avoid some of the over-generalization of large-N statistical studies and the over-specificity of case studies. Likewise, the comparative method can avoid many of the problems of external validity that are inherent in experimental research designs.

Given those general virtues of the comparative method, are there specific benefits that this method can bring to the study of public policy? Without making excessive claims, I would argue that there are. Much of the public policy literature, especially that based in political science, focuses on the processes of making policy (see Peters and Zittoun 2016).

Understanding the policy process well involves more detailed tracking of events in something like case analysis (see above), and more specifically process tracing.

Therefore, if the researcher wants to understand variations among processes and the consequences of those variations then the comparative method may be the preferred analytic approach. But again that emphasis on comparison may well be allied with other approaches and techniques that can be used to identify the evidence that will feed into the comparison. Once the evidence is gathered then it can be fed into the analysis across cases that should yield the comparative insights.

The comparative method also attempts to bring context more directly into an analysis than do quantitative and experimental methods. Indeed, the experimental method attempts to exclude context as extraneous noise in the relationships being observed. For the study of public policy in many situations context is crucial for understanding the dynamics involved. Policy making is not taking place in some sterile laboratory environment but rather in complex institutions which have values and routines (March and Olsen 1989) that influence the outcomes of the process as much or more than individual preferences. But, as already noted, there is a tension between bringing in context (again, as in case studies) and attempting to control context for more scientific validity.

Following from the importance of recognizing the importance of context in the comparative method, this approach to comparative policy requires some justification of case selection. That sounds rather fundamental to any research, but it is remarkable how many studies either take all available cases with little concern for their relevance (see Della Porta and Keating 2008), or select cases on the basis of convenience. The comparative method requires some prior knowledge and some logical justification for the inclusion of a case in the analysis.⁸ The logic therefore is very different from that of statistical analysis, in which a larger sample improves, *ceteris paribus*, the possibilities of generalization.

The comparative method, as outlined by Lijphart, provides many opportunities for scholars of public policy to enhance our collective understanding of policies and the policy process. Using this method may not be, however, as easy to apply as other methods commonly used in the social sciences. Utilizing the comparative method requires substantial knowledge of the cases that are available before making the selection. The comparative method also requires detailed understanding of the cases that are selected, especially if methods such as process tracing are employed in the analysis of the cases. And the comparative study of public policy also involves having relatively well-articulated theories that can inform the researcher about what variables are relevant and which not when she or he selects the cases.

Further, in addition to the knowledge requirements for engaging in the comparative method effectively, this method may have difficulty in coping with some of the more challenging aspects of comparative politics and public policy. The most challenging one for policy may be complexity (Cairney and Geyer 2015) and the non-linear relationships among variables that produce final policy results. Associated with that complexity are the problems of equifinality and multifinality. Although comparative policy analysis should examine process, one of the standard tenets of complexity is that many routes can produce the same final results, and the same initial conditions can produce different outcomes.

The study of complexity therefore introduces greater uncertainty into the study of public policy, and also makes humility more necessary for the practitioners of this craft. Science demands replicability of outcomes, but one of the principles of complexity is that any one set of outcomes may not be replicable. This dilemma, along with the knowledge demands, make

doing comparative policy analysis well a daunting task. However, if those conditions can be met then these comparisons can yield very rich insights into the causes of policy choices, as well as their consequences.

NOTES

1. This work was originally presented at the Workshop on Comparative Methodology at IPPW, University of Pittsburgh, Pittsburgh, PA, June 26–28, 2018.
2. This form of error is usually assumed away in quantitative research by assuming that across all observations the mean of error is zero. And we have to assume that the “hard” quantitative evidence is the product of human collection and processing, so may contain the same sort of errors as does qualitative evidence.
3. Even then, however, there can be other variables lurking that have not been introduced that are the real cause of the relationship. However, with quantitative research it is relatively easy to re-run the model with a new variable introduced. Indeed, much of the quantitative research published seems to be rather simplistic additions of one or more variables to familiar models.
4. The contemporary fad of natural experiments tends to violate this assumption and hence raises significant questions of the internal validity of the findings, along with the questions of external validity and replication that bedevil the more controlled experimental research. See Sekhorn and Titunik (2012).
5. In this case it is important to think about the nature of policy and policy problems in more analytic terms as well as the usual functional categories such as “health” or “defense”. See Hoornbeek and Peters (2016).
6. If a variable is present in one case of success and absent in another it cannot be a necessary condition for the success. It may still be sufficient in some situations, but again there is the need to contextualize any findings.
7. Several iterations of this process are possible, with Lieberman providing some advice about when to terminate the analysis. Note also that this version of the linkage between quantitative and qualitative analysis assumes that the goal is to have a better quantitative model, not to have the understanding of dynamics that can come from the case, or comparative, analysis.
8. This does not mean that the justification for the selection of most-similar cases will necessarily be correct. How many studies are there that conceptualize the United States and the United Kingdom as most similar when in reality they are fundamentally different in most important respects?

REFERENCES

- Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science* 59: 495–510.
- Allardt, E. (1990). Challenges for comparative social research. *Acta Sociologica* 33: 183–93.
- Beach, D. and Rohlfing, I. (2018). Integrating cross-case analysis and process-tracing in set theoretic research: strategies and parameters for debate. *Sociological Methods & Research* 47 (1): 3–36.
- Berg-Schlosser, D., De Muer, G., Rihoux, B., and Ragin, C. C. (2009). Qualitative Comparative Analysis (QCA) as an approach. In B. Rihoux and C. C. Ragin (Eds.), *Configurational Comparative Methods*. London: Sage, 1–18.
- Blatter, J. and Blume, T. (2008). In search of co-variance, causal mechanisms or congruence: towards a plural understanding of case studies. *Swiss Political Science Review* 14 (2008): 315–54.
- Blatter, J. and Haverland, M. (2014). Case studies and (causal-) process tracing. In I. Engeli and C. Rothmayr Allison (Eds.), *Comparative Policy Studies: Conceptual and Methodological Challenges*. Basingstoke: Palgrave Macmillan, 59–84.
- Cairney, P. (2016). *The Politics of Evidence-Based Policy Making*. London: Palgrave Macmillan.

- Cairney, P. and Geyer, R. (2015). Introduction. In R. Geyer and P. Cairney (Eds.), *Handbook on Complexity and Public Policy*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing, 1–18.
- De Winter, L., Della Porta, D., and Deschouwer, K. (1996). Comparing similar countries: Italy and Belgium. *Res Publica* 38: 215–35.
- Della Porta, D. and Keating, M. (2008). How many approaches in the social sciences? An epistemological introduction. In D. Della Porta and M. Keating (Eds.), *Approaches and Methodologies in the Social Sciences: A Pluralist Perspective*. Cambridge: Cambridge University Press, 19–39.
- Dunning, T. (2012). *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge: Cambridge University Press.
- Falletti, T. G. and Lynch, J. F. (2009). Context and causal mechanisms in political analysis. *Comparative Political Studies* 42 (9): 1143–66.
- Freeman, G. P. (1985). National styles and policy sectors: explaining structured variation. *Journal of Public Policy* 5 (4): 467–96.
- Geddes, B. (2003). *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics*. Ann Arbor, MI: University of Michigan Press.
- Gerring, J. (1999). What makes a concept good? A criterial framework for understanding concept formation in the social sciences. *Polity* 31: 351–93.
- Gerring, J. (2007). Is there a (viable) crucial case method? *Comparative Political Studies* 40: 231–53.
- Hoornbeek, J. A. and Peters, B. G. (2016). Understanding policy problems: a refinement of past work. *Policy & Society* 36: 365–84.
- Lieberman, E. S. (2005). Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review* 99 (3): 435–52.
- Lijphart, A. (1971). Comparative politics and the comparative method. *American Political Science Review* 65: 682–93.
- Mahoney, J. (2007). Qualitative methodology and comparative politics. *Comparative Political Studies* 40: 122–44.
- March, J. G. and Olsen, J. P. (1989). *Rediscovering Institutions: The Organizational Basis of Political Life*. New York: Free Press.
- Morlino, L. (2018). *Comparison: A Methodological Introduction to the Social Sciences*. Opladen: Barbara Budrich.
- Peters, B. G. (2013). *Strategies for Comparative Research in Political Science*. Basingstoke: Palgrave Macmillan.
- Peters, B. G. and Zittoun, P. (2016). *Contemporary Approaches to Public Policy*. London: Palgrave Macmillan.
- Pollitt, C. (2013). *Context in Public Policy and Management: The Missing Link?* Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Przeworski, A. and Teune, H. (1970). *The Logic of Comparative Social Inquiry*. New York: John Wiley.
- Sartori, G. (1970). Concept misformation in comparative research. *American Political Science Review* 64: 1033–53.
- Scruggs, L. and Allan, J. (2006). Welfare-state decommodification in 19 OECD countries. *Journal of European Social Policy* 15: 55–72.
- Seawright, J. and Gerring, J. (2008). Case selection techniques in case study research: a menu of quantitative and qualitative options. *Political Research Quarterly* 61: 294–308.
- Sekhorn, J. S. and Titunik, R. (2012). When natural experiments are neither natural nor experiments. *American Political Science Review* 106: 35–57.
- Toshkov, D. (2018). Small-N comparative research: a viable strategy for explanatory analysis in the field of public policy. Paper presented at Workshops of the International Public Policy Association, Pittsburgh, PA.
- van den Heijden, J. (2014). Selecting cases and inferential types in comparative public policy research. In I. Engeli and C. Rothmayr Allison (Eds.), *Comparative Policy Studies: Conceptual and Methodological Challenges*. Basingstoke: Palgrave Macmillan, 35–56.
- Verhoest, K., van Thiel, S., Bouckaert, G., and Laegreid, P. (Eds.) (2012). *Government Agencies: Practices and Lessons from 30 Countries*. Basingstoke: Palgrave Macmillan.

Yom, S. (2015). From methodology to practice: inductive iteration in comparative research. *Comparative Political Studies* 48 (5): 616–44.

3. The most-similar and most-different systems design in comparative policy analysis

*Carsten Anckar*¹

1. INTRODUCTION

Five decades ago, Arend Lijphart (1971), in an often quoted article, distinguished between four different methods that were available in political research, namely *the experimental method*, *the statistical method*, *the comparative method*, and *the case study*. He concluded that ‘[t]he experimental method [was] ... the most nearly ideal method for scientific explanation, but it [could] ... only rarely be used in political science because of practical and ethical impediments’ (Lijphart 1971: 683–4). Instead, researchers who wanted to come up with generalizations on social phenomena were referred to the statistical method and the comparative method.

To a certain extent, Lijphart’s conclusions and recommendations are still valid but in a number of respects the article has not withstood the test of time. First, experimental methods are used much more in political science today than only a decade ago. At the same time, we cannot escape the fact that its level of application is largely restricted to areas where the phenomenon of interest resides at the individual level. In comparison with natural scientists, experimental researchers in political science are also confronted with disadvantages such as an artificial environment, unrepresentative subject pools, external validity, and experimenter bias (McDermott 2002: 39–40).

The statistical method, again, has strengthened its position in comparative research. This is hardly surprising given the fact the number of large data sets has increased substantially. Often these data sets are expansive in time and space and, importantly, freely available for public download (e.g. Anckar and Fredriksson 2019; Coppedge et al. 2018). The other reason for the growing popularity of the statistical method is the entry of highly sophisticated software packages, which make it possible to conduct advanced quantitative analyses fairly easily.

Although both the experimental method and the statistical method have gained popularity, case studies and small-N studies continue to be influential in the social sciences (Gerring 2017: 8–10; Herron and Quinn 2016: 459). Many of these studies follow the logic of what Arend Lijphart, in his article referred to as the comparative method. However, Lijphart’s conception of the comparative method is hardly compatible with the connotation the concept has today. Instead, what Lijphart referred to as the comparative method comes very close to what nowadays generally is referred to as the *most-similar systems design*.

2. THE MOST-SIMILAR SYSTEMS DESIGN

For Lijphart (1971: 684) the logic of the comparative method was essentially the same as the logic of the statistical method: ‘The comparative method resembles the statistical method in all respects except one. The crucial difference is that the number of cases it deals with is too small

to permit systematic control by means of partial correlations'. A few years later he clarified this statement and defined the comparative method as 'the method of testing hypothesized empirical relationships among variables on the basis of the same logic that guides the statistical method, but in which the cases are selected in such a way as to maximize the variance of the independent variables and to minimize the variance of the control variables' (Lijphart 1975: 164).

Today, we do not refer to this research strategy as *the* comparative method. Instead authors most often use the term *the most-similar systems design*, a concept which had been introduced by Przeworski and Teune (1970) a year before Lijphart's seminal article was published. By applying this research strategy we choose, as objects of research, systems that are as similar as possible, except with regard to the research phenomenon of interest. The reason for choosing systems that are similar is an ambition to keep constant as many extraneous variables as possible (e.g. Bartolini 1993: 134; Sartori 1991: 250; Skocpol 1984: 379). The method follows the logic of John Stuart Mill's (1970 [1843]: 256) method of difference, which states that '[i]f an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon'.

There is one important difference between Lijphart's conception of the comparative method and Mill's method of difference. Lijphart explicitly argued that when the researcher chooses objects of research, the ambition should be to maximize variation on the independent variable and minimize variation on the control variables. However, the dependent variable should be left out of consideration when choosing objects of research as 'this would prejudice the empirical question' (Lijphart 1975: 164). Mill, on the other hand, did not stipulate that we should begin the research process by securing variation on the independent variable. In other words, whereas Lijphart restricted the application of his *comparative method* to studies that follow a strictly deductive approach, Mill and most other scholars accept, explicitly or implicitly, that it can be used also in inductive studies.

3. THE MOST-DIFFERENT SYSTEMS DESIGN

The main disadvantage with the most-similar systems design is that the limited number of countries in the world makes it impossible to find countries that are similar in all relevant aspects except for the central phenomenon under investigation. As a solution to this problem, Przeworski and Teune (1970) introduced another, opposite, research strategy, which they labelled the *most-different systems design*. Instead of controlling for extraneous variance, the strategy is to eliminate as many external variables as possible from the analysis. This objective is achieved by choosing units of analysis, which are as different as possible with regard to extraneous variables. Echoing Popper (1959), the claim is that falsification rather than verification is central for the progress of science.

When we make use of the most-different systems design we test a specific proposition in sub-systemic settings. The system level enters the research design only if, and when, the analyses within systems show that different variable associations exist within different systems (Przeworski and Teune 1970: 35). If a relationship between two variables, X and Y, is detected in countries A, B and C, but not in D, the assumption is made that there is some systemic

quality, Z, which distinguishes D from the other countries. This quality, or systemic variable, affects the relationship between X and Y.

Although the most-different systems design handles the ‘many variables, small-N problem’ better than the most-similar systems design, it apparently suffers from another drawback, namely that, if Przeworski and Teune are to be believed, it can only be applied in situations where the dependent variable resides at a sub-systemic level. However, this qualification has been challenged, explicitly and implicitly, by a number of authors. Sartori (1991: 250), for instance, noted that the requirement that the most-different systems design must operate at a sub-systemic level ‘is a differentiation open to question’ (see also Frensdreis 1983). Indeed, when authors want to exemplify how the most-different systems design can be applied in empirical studies, a recurring example is Theda Skocpol’s *States and Social Revolutions* (1979). In Skocpol’s work the dependent variable was the occurrence of major revolutions, which, naturally, resided at the systemic level.

4. DEDUCTIVE AND INDUCTIVE RESEARCH STRATEGIES

There are two basic strategies available for a researcher who wants to arrive at generalizations: deduction or induction. When using a deductive approach our point of departure is a theory, or a theoretical proposition, which we set out to test. This means that we have a notion of a causal link that exists between our *explanans* (our independent variable) and our *explanandum* (our dependent variable). When we use a deductive approach the research question is expressed in the sentence ‘Does X affect Y?’. We are, in other words, not primarily concerned with explaining all the variation in the dependent variable, i.e. accounting for the explanatory value of every single explanatory factor. Instead, we have the ambition to establish whether or not there is a causal relation between one specific independent variable and the dependent variable.

When we apply an inductive strategy there is no established theory which could be used as a point of departure. Instead, focus is on the dependent variable and the research question is consequently formulated in the sentence ‘What explains Y?’, indicating that the ambition is to ‘discover’ the relevant independent variable(s). In practice, the difference between the two strategies is likely to become blurred. It is difficult to imagine a purely inductive study in the social sciences. Although there is no well-formulated theory which explains a certain phenomenon, we generally know at least in which settings it makes sense to look for plausible determinants of the phenomenon in question. Furthermore, a ‘what explains y-question’ could also be answered by means of deductive reasoning if the researcher has the ambition to identify (nearly) all plausible determinants of the dependent variable by means of theoretical reasoning.

Another example of the difficulties involved in separating deduction from induction is the standard ordinary least squares regression analysis. Most often, the dependent variable is related to a specific independent variable, which has been chosen based on theoretical reasoning. At the same time the researcher includes a number of control variables, but the theoretical justification for the inclusion of these variables is often very vague (as we know, it is not uncommon that the only motivation given for the inclusion of some control variables is that they constitute ‘usual suspects’). It should be noted, though, that in a standard ordinary least squares regression, the position of the key independent variable is not different from the position of the control variables since we measure the effects of each independent variable on

the dependent variable at constant values on all other independent variables. In other words, we cannot assess the impact of one independent variable without assessing the impact of all other independent variables included in the regression model.

5. APPLYING THE MOST-SIMILAR SYSTEMS DESIGN IN DEDUCTIVE AND INDUCTIVE RESEARCH

When applying the most-similar system design, the distinction between deduction and induction is important because it has quite profound consequences for how the method is applied, especially with regard to the selection of cases. Contrary to what is the case in multiple regression analysis the position of the independent variable differs from the position of the control variable(s) when we make use of a most-similar systems design. In a deductive study, where the research question is framed in a ‘Does x explain y-sentence’, we would follow Lijphart’s advice and choose cases with maximal variation on the independent variable and minimal variation on the control variables. Values on the dependent variable, however, should not be considered when selecting the cases (Gerring 2017: 95–8; King, Keohane, and Verba 1994: 137–41; Seawright and Gerring 2008: 304–5).

The logic of this design is straightforward and theoretically robust. Its shortcoming is of a practical nature. A researcher who tries to apply a most-similar systems design is often confronted with what Lijphart (1975: 163) referred to as *the many variables, small N-problem*. Since there is a limited number of countries, it is very difficult to keep constant all potential explanatory factors (e.g. Meckstroth 1975: 134; Peters 1998: 38–9). A practical consequence is that it is not easy to find studies where the most-similar systems design is applied in its strictest sense, i.e. a design where the countries chosen for the study are similar in a number of clearly specified variables (the control variables) and different with regard to only one aspect (the independent variable). More common is therefore to conceive of the most-similar systems design in a broader sense. This happens when a researcher chooses countries that appear to be similar in as many background characteristics as possible, but never systematically matches the cases on a specific set of control variables. In Della Porta’s (2008: 214) words, by means of such a strategy ‘many variables are “parametrized”’: if we have more or less the same degree of economic development, similar culture and the like, we can consider these characteristics as constant and check for the influence of other factors’. If the most-similar systems design is conceived of in the latter form, most comparative studies confined to specific geographic settings could be said to implicitly apply a most-similar systems design.

If the research follows an inductive approach, the research question is framed in a ‘What explains y-sentence’ and our task is to find the relevant independent variable(s) of the phenomenon of interest. In such situations it would be natural to choose cases which vary with regard to the dependent variable (compare with Gerring 2017: 40–41, 63–98). However, when induction is used in its purest form we do not know which extraneous variables to keep constant. Therefore, the strategy is instead to try to keep as many plausible extraneous variables as possible constant by choosing countries that are (or appear to be) similar with regard to as many features as possible (e.g. geography, history, culture, development, etc.). These similarities cannot explain the different outcomes on the dependent variable, and can consequently be excluded as plausible explanations of the dependent variable. The goal is therefore to identify a plausible explanatory variable which varies according to the same pattern observed on the

dependent variable (Ragin 1987: 47). For excellent illustrations of how such a design can be applied empirically, the reader is referred to Berg-Schlusser and De Meur (1994) and De Meur and Berg-Schlusser (1996).

6. APPLYING THE MOST-DIFFERENT SYSTEMS DESIGN IN DEDUCTIVE AND INDUCTIVE RESEARCH

Like the most-similar systems design the most-different systems can be used in both deductive and inductive research designs. When the most-different systems design is applied in a purely deductive ('Does x affect y?') study we test a specific hypothesis in several contexts which differ maximally from each other. Przeworski and Teune (1970: 35) provided the following example: 'If education is positively related to attitudes of internationalism in India, Ireland and Italy, the differences among these countries are unimportant in explaining internationalist attitudes'. In other words, since differences cannot explain similarities, all conditions on which the contexts differ from each other can be eliminated from the analysis.

At the systemic level it is possible to apply the most-different systems design in both deductive and inductive studies. When we apply it in inductive studies ('What explains y?'), the natural thing is to operate with a constant dependent variable in systems that differ maximally from each other. The ambition would then be to identify a common feature (the relevant explanatory variable) in all these different systems. If the research question is deductive ('Does x affect y?') it makes sense to focus instead on the relevant independent variable. In such cases, we would choose countries that differ maximally from each other except with regard to the independent variable. We would thereafter study if the dependent variable has the same value in all these countries. Another variant is to select cases which have the same value on the dependent variable and then study if the expected independent variable also has the same value in all cases. In both forms, the ambition is to match similarities in different systems, and the method is identical to Mill's (1970 [1843]: 255) 'method of agreement', which states that '[i]f two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all of the instances agree is the cause (or effect) of the given phenomenon'.

There is, however, an important difference between these procedures. If we start the process by selecting cases that have the same value on the independent variable of interest, and then note that all of these cases also have the same value on the dependent variables, we reach the conclusion that X is a *sufficient* condition of Y. In other words, based on our findings we know that when X is present, Y is always present. However, we do not know if it is possible for Y to occur without the presence of X (since our population does not include countries where X is absent). This question is instead answered if we operate with an inductive approach and start by selecting cases that have the same value on the dependent variable. We would then, perhaps, reach a result showing that in all cases where Y occurs, X is also present. In such a situation X is a *necessary* condition of Y. However, we do not know if X always generates Y (since we have not included cases where Y is absent). Therefore, we do not know if X is a sufficient condition of Y (Dion 1998).

7. THE MOST-SIMILAR AND MOST-DIFFERENT SYSTEMS DESIGNS COMPARED

In contrast to the method of agreement, the method of difference allows us to reach conclusions about both necessity and sufficiency. This is so, because regardless of whether we choose cases based on their values on the independent or on the dependent variable, we always secure variation on the phenomenon of interest. In other words, our research design allows us to answer if X always gives rise to Y, and, also, if Y can occur without the presence of X. The aforesaid means that when applied at the systemic level, the most-similar systems design is a better research strategy than the most-different systems design, as it is possible to find out if an independent variable is a necessary or sufficient (or both) condition of the dependent variable.

Although the most-similar systems design and the most-different system design differ from each other in many respects, the system level occupies an important role in both designs. This is the case also when the most-different systems design is applied the way Przeworski and Teune suggested, i.e. for studying variable interactions in sub-systemic settings. The reason for this is that patterns between independent and dependent variables become particularly interesting when they are confirmed (or disconfirmed) in different contextual settings.

It should be acknowledged that the logic of the most-similar and most-different systems designs can be applied also in situations when we are exclusively interested in variable relations at the lowest level of analysis. Thus, at least in principle, it is possible to study if income is related to negative attitudes towards government spending by choosing individuals which differ with regard to income but show a resemblance on a number of other plausible determinants of attitudes towards government spending. However, since the number of cases can be extended quite easily when we study characteristics at the individual level, we would instead prefer to use more sophisticated statistical methods for measuring the association between the variables in question.

The multilevel approach is often regarded as a distinguishing feature of the most-different systems design, but as we have seen, it is probably fair to say that a vast majority of authors are of the opinion that the most-different systems design can perfectly well also be used at the system level. Similarly, the multilevel approach associated with Przeworski and Teune's most-different systems design is applicable also in studies which follow a most-similar systems design (Frendreis 1983: 260; Przeworski and Teune 1970: 33–4). Suppose, for instance, that we expect a variable operating at the system level to influence how an independent and a dependent variable are related to each other at the sub-system level. We would then choose, say two, countries that are extremely similar, except with regard to the system-level variable of interest, and then study if the relationship between the independent and dependent variable varies between the countries.

It should also be stated that a researcher is not necessarily confronted with the choice of applying either a most-similar systems design or a most-different systems design. In fact, the two strategies can often be combined within the framework of a single research effort (Denk 2010; Faure 1994; Frendreis 1983). Sometimes this is done explicitly and sometimes more implicitly (e.g. Collier and Collier 1991; Linz and Stepan 1996). A classic example of an explicit attempt to combine the two designs is the studies of the conditions for survival and breakdown of democracy in interwar Europe undertaken by De Meur and Berg-Schlosser (Berg-Schlosser and De Meur 1994; De Meur and Berg-Schlosser 1994, 1996). In their studies

they combined both Mill's method of agreement (which they labelled *most different with same outcome* (MDSO)) and Mill's method of difference (which they referred to as *most similar with different outcome* (MSDO)).

8. CRITICISM AND RESPONSE

The most-similar and most-different systems designs are based on Mill's methods of difference or agreement. The feasibility of these methods has been debated quite extensively in the political science literature. One point of criticism is that the methods are deterministic, and thus unable to cope with probabilistic assumptions and measurement errors (Liebersohn 1991). The counterargument is that if we make the claim that no exception from a rule is allowed when applying Mill's methods, it means that Mill's methods and regression analyses are judged by double standards. A single case which contradicts a pattern in regression analysis does not mean that the hypothesis is rejected, and thereby one can ask oneself why this should be the case in a most-similar systems design where one single case contradicts the hypothesis (compare with Goertz 2005).

Another frequently occurring criticism against the most-similar and most-different systems designs (or, rather Mill's methods of difference and agreement) is that they are unable to cope with interaction effects and multiple causation (Liebersohn 1991, 1994; and for a rebuttal Savolainen 1994). In an earlier contribution (Anckar 2008) I argued that Mill's methods can, indeed, be used both for creating interaction terms as well as for discovering interaction effects in the empirical material. The argument is repeated here by using an example with more relevance for comparative public policy.

Suppose that our research task is to explain why some countries make use of the death penalty and some countries do not. We then set out to test the hypothesis that a democratic form of government is related to the abolition of the death penalty. This assumption is based on the argument that a democracy is built on the conviction that all humans have inalienable rights, most importantly, the right to life. We therefore expect that the principle of human rights influences how penal codes are shaped (Anckar 2014: 11–12; Brettschneider 2002; Burt 1994: 90; Neumayer 2008a: 6). We apply a most-similar systems design for this deductive research question ('Does democracy explain the abolition of the death penalty?'). Accordingly, we should choose countries which differ with regard to the independent variable, i.e. regime type, but are similar in other respects.

We also introduce two control variables, population size and a British colonial heritage. Both of these variables can be theoretically linked to death penalty usage. Populous countries are stronger than smaller countries and thereby more likely to resist international pressures for abolition of the death penalty (Neumayer 2008a: 12). Also, as small entities are characterized by intimacy and dependency, the thought of punishing criminals by taking their life can be considered repulsive (Anckar 2004: 25). A British colonial heritage can also affect attitudes towards the death penalty. Neumayer (2008a: 11) has pointed out that '[w]ith the spread of the British Empire the common law legal system and the death penalty came to be applied in its colonies'. Accordingly, it has been suggested that the popularity of the death penalty in the English-speaking countries in the Caribbean is attributed to the fact that the countries in question inherited the death penalty from their former colonizer and continued using it even after Britain became an advocate of the abolitionist movement (Knowles 2004).

Table 3.1 *Coping with interaction effects in a most-similar systems design*

	São Tomé and Príncipe	Comoros	Poland	Belarus	South Africa	Zimbabwe	Vanuatu	Singapore
Democracy	Yes	No	Yes	No	Yes	No	Yes	No
Small population	Yes	Yes	No	No	No	No	Yes	Yes
British colonial heritage	No	No	No	No	Yes	Yes	Yes	Yes
Death penalty abolished	Yes	No	Yes	No	Yes	No	Yes	No

When operationalizing the independent variables we regard countries rated ‘free’ by Freedom House² as democracies and all other countries non-democracies. The size variable is dichotomized; the threshold for inclusion among the large countries is set at 10 million inhabitants.

Since our approach is deductive, and we wish to test if regime type is related to death penalty status we secure variation on the independent variable of interest while keeping extraneous variables constant. In Table 3.1, we have started by choosing two countries, São Tomé and Príncipe and Comoros, which differ with regard to regime type but are similar with respect to the two control variables. We also note that São Tomé and Príncipe has abolished the death penalty whereas Comoros continues to make use of it. In other words, since similarities cannot explain differences our conclusion is that the hypothesis is confirmed. Such a conclusion overlooks the possibility that interaction effects between the variables might be of relevance for the outcome. It might well be that the real explanation for the absence of death penalty in São Tomé and Príncipe is a combination of democracy and a small population size, or a combination of democracy and a lack of British colonial heritage, or a combination of democracy, a small population size, and a lack of British colonial heritage.

It is, however, possible to control whether interaction effects operate in the example above. However, in order to do so we must include more cases into the research design. We therefore continue by introducing Poland and Belarus into the analysis. Thereby we can rule out the combination democracy/small size as an explanation of death penalty abolition. Moving on, we introduce the pair South Africa and Zimbabwe in the analysis, whereby the combination democracy/lack of British colonial heritage is falsified. Finally, when Vanuatu and Singapore are included we can rule out the possibility that democracy in combination with *either* small population or a lack of British colonial heritage is required for an absence of death penalty in the criminal code.

The example above shows that at least in theory it is possible to account for interaction effects when using Mill’s methods. This presupposes a careful selection of cases. When selecting the cases, it is important to maximize the variation in the phenomenon of interest, while keeping control variables constant. However, since it is possible that the dependent variable is affected by other variables than the ones we explicitly control, a guiding principle is to choose the cases with an ambition to implicitly control for as many other explanatory variables as possible. In the example above, we have chosen pairs of countries that are confined to the same geographical region and are either island states or mainland states. This is more or less a ‘standard operating procedure’ when applying the most-similar systems design. However, it is

not a rule without exceptions. For instance, in a number of respects Australia and New Zealand have much more in common with Britain and Canada than with their neighbours Papua New Guinea or Indonesia. Also, the former British colonies in the Pacific share many important features with former British colonies in the Caribbean on the other side of the globe.

If studies are deductive and follow the logic of method of difference, interaction effects can be controlled in the way exemplified above. Interaction effects can also be tested by applying the method of agreement in deductive studies. Suppose, for instance that we expect X1 to affect Y and test this proposition in as varying contexts as possible. However, the results show that the occurrence of X1 does not always lead to outcome Y. However, we then consider if the combination of X1 and another plausible determinant of Y, variable X2, is crucial for Y to occur. If all cases where variables X1 and X2 occur together render the outcome Y, we reach the conclusion that the interaction between X1 and X2 is sufficient for the outcome.

The problem of handling interaction effects is more complicated in inductive studies since we have no theoretical a priori notion of which combinations we should try to isolate. However, if we have a situation where we find that two or more countries have similar values on the dependent variable but also on two or more plausible independent variables, the question whether or not the respective variables are capable of generating the outcome on their own or only in combination with each other can be solved by including more cases in the analysis.

Mill's methods have also been criticized for not being able to handle the problem of multiple causation (Lieberman 1991, 1994; Thiem 2014: 20), or situations where different independent variables generate an outcome. However, if we apply the logic of method of difference in deductive studies, the problem can be remedied in the same way as in Table 3.1, i.e. by including more cases in the analysis in order to let the independent variables vary while keeping extraneous variables constant.

However, if we apply the methods in inductive research projects the argument is valid. Suppose we make use of the method of agreement and select cases which are constant on the dependent variable. If both X1 and X2 generate Y, some instances of Y have X1 in common and some instances X2 and none of the independent variables are perfectly associated with the outcome. Similarly, in a most-similar systems design in which we would start by securing variation on the dependent variable, instances of X1 would have Y in common. However, since this would also be the case for instances with X2, no relationships between any of the independent variables and dependent variables would be detected.

It should also be readily admitted that even a careful selection of cases does not mean that we are able to control for all possible extraneous variance (e.g. Ebbinghaus 2005: 142–3). In essence, it is very difficult to overcome the 'many variables, small N-problem' as the limited number of countries in the world places restrictions on the extent to which we can include new countries in the research design in the manner described in Table 3.1. In the example described above, our ambition was to explain why some countries make use of the death penalty whereas other countries do not. Suppose, now, that we set out to test the same research question with a most-different systems design, and choose countries that are maximally different except with regard to the research phenomenon of interest. Suppose further that we come up with a result which shows that all countries which make use of the death penalty are non-democratic. Since we have chosen a most-different systems design, the countries included in the study vary in a number of characteristics that could constitute plausible determinants of the use of the death penalty. In addition to British colonial heritage and population size such explanatory variables could be *level of crime, history of slavery, dominating religion, socio-economic development,*

level of corruption, conflict intensity, ideological orientation of government, ethnic fractionalization, etc. (Anckar 2004; Neumayer 2008b). Let us further suppose that an examination of the cases reveals that none of the authoritarian countries making use of the death penalty exhibited a high level of socio-economic development, no history of slavery, a low level of corruption, a low level of conflict intensity, Taoism as dominant religion, a left-wing executive and a low level of ethnic fractionalization. Consequently, we do not know if this particular combination also would generate a state which makes use of capital punishment or not.³

9. THE RELEVANCE OF THE MOST-SIMILAR AND MOST-DIFFERENT SYSTEMS IN CONTEMPORARY RESEARCH

It is noteworthy that most of the pioneering works on the methods were written many decades ago. It is therefore relevant to ask if the most-similar systems design and the most-different systems design have become obsolete? Two recent methodological developments in particular have challenged the utility of the most-similar and most-different systems designs in social research: *qualitative comparative analysis* and *multilevel modelling*.

9.1 Qualitative Comparative Analysis

We cannot escape the fact that Mill's methods suffer from shortcomings when applied at the system level. The fact that the number of countries is limited often makes it difficult to systematically compare different sets of countries in order to cope with interaction effects and multiple causation in the manner described under the previous heading. In many ways it is natural to regard qualitative comparative analysis (QCA) as an ameliorated variant of the most-similar and most-different systems designs. Indeed, it is telling that when Charles Ragin (1987) introduced the method into the social sciences, he referred to it as *the comparative method*, in other words he used the same term Lijphart had used for the most-similar systems design in the early 1970s. In many ways the QCA can be seen as an extension of Mill's methods of difference and agreement (Hug 2013; Ragin 1987: 36–42), although it should be readily admitted that some authors disagree (Thiem 2014: 20).

The QCA approach and its applicability for comparative public policy is discussed extensively in other parts of this volume and I shall not dwell on the subject. However, some points are noteworthy. In short, QCA can be construed as '[a] systematization of small-N comparative analysis and analytic induction' (Seawright and Collier 2010: 344). It is worth pointing out that QCA copes better than Mill's methods with interaction effects and multiple causation. The point of departure for the QCA approach is that it focuses on whether or not a condition (independent variable) is necessary or sufficient for an outcome (the dependent variable) and that an outcome can be achieved by many different configurations, or combinations of independent variables (Ragin 2008; Schneider and Wagemann 2012). In this respect QCA shows similarities with the most-similar and most-different systems design, as these methods too, base their logic on the concepts of necessity and sufficiency.

It should, however, also be stated that throughout the years Ragin's QCA approach has been developed extensively and now encompasses a wide variety of sophisticated techniques (Cronqvist and Berg-Schlosser 2009; Grofman and Schneider 2009; Mahoney and Vanderpoel

2015; Ragin 2000, 2008; Rihoux and Ragin 2009; Rohlfing 2016). In many respects, the differences between QCA and the statistical methods have become less evident. At the same time its utility continues to be questioned and there is an ongoing debate between proponents of the method and advocates of regression analysis (e.g. Goertz and Mahoney 2012; Paine 2016; Thiem and Baumgartner 2016).

Although the QCA approach is a useful tool for assessing multiple causation in small-N analyses, there are some differences between its logic and that of the most-similar and most-different systems designs. As we have seen, both the most-similar systems design and the most-different systems design follow a variable-oriented approach. The QCA technique, on the other hand, takes a different approach. The aim is not primarily to test a hypothesis but rather to construct and refine theories by means of an ongoing interplay between theory and data. In other words, whereas the most-similar systems design in particular often aims at *testing* theories, the primary goal of using QCA is to *discover* empirical relationships that can help us formulate theoretical proposals. Researchers who make use of QCA do not want to test how strong the relationship between an independent and a dependent variable is, but instead argue that an outcome (a certain value on the dependent variable) can be reached by several combinations of conditions (i.e. independent variables). The QCA method inevitably has a lot in common with the most-similar and most-different systems designs, but this is more the case when the methods are used for inductive purposes than when they are used in deductive research efforts.

Although the introduction of QCA has challenged the role of the most-similar and most-different systems techniques, the methods can still play an important role in social research. By applying the most-similar and most-different systems designs, it is possible to exclude contesting variables from the analysis by carefully matching the cases according to the principles of the two methods. The central feature when applying the most-similar systems design and the most-different systems design is the ambition to isolate the explanatory value of the independent variable as much as possible. This is done by choosing countries that are as similar as possible on the background variables (in a most-similar systems design) or as different as possible (in a most-different systems design).

In addition, the most-similar and most-different systems designs fall nicely in between the statistical method and case studies. The number of cases is lower than in a statistical study but larger than in a case study. This means that they share the variable-oriented approach with the statistical method, but allow the researcher to account also for complex social processes that are difficult to reduce to a simple point in a data set. An example could be if we aim to explain if, when, or how certain policy proposals are initiated and whether they become adopted or not. Such processes are often complex and require substantial knowledge about the countries included in the study.

9.2 Multilevel Analysis

It is now half a century since Przeworski and Teune introduced their most-different systems design. During these years a lot of methodological advances have been achieved in the social sciences. The one with the largest impact on the most-different systems design, at least in the sense it was conceived of by Przeworski and Teune, is multilevel analysis.

As we recall, Przeworski and Teune applied the most-different systems design in deductive research efforts in which the dependent variable resided at the sub-systemic level. We first

study relationships between variables within different systems and then compare the results obtained between the systems. If the results differ between the systems, we turn to the system level (i.e. we introduce independent variables that denote system-level characteristics). The advantage of multilevel modelling is that it makes it easier to shift the level of analysis. We no longer have to exhaust the variables at the sub-system level before turning to the next level of analysis. Instead, we can study effects of variables at both analytical levels at the same time. When Przeworski and Teune published their book it was not possible to conduct multilevel analyses, although it should be clearly stated that such a statistical tool was precisely what Przeworski and Teune (1970: 72) desired: '[w]hat we need in comparative research ... are statistical techniques that would allow the control variable to be measured at a level different from the two variables that are tested'. Today such techniques are, indeed, available but this does not mean that we should reject the most-different systems design as obsolete.

The reason for this is, once again, the perennial problem all researchers engaged in cross-country comparisons regularly face, namely the fact that the number of cases is restricted to the 200 or so independent countries of the world. This means that if we start introducing variables residing at the systemic (country) level into the model, we very soon have to confront the problem of multicollinearity at the systemic level. Consider a researcher who wants to explain attitudes towards abortion with regard to the level of education. The researcher additionally believes that system-level characteristics such as democracy, secular state, left-wing government, and high level of socio-economic development are important. The researcher in question will find it hard to build robust statistical models since values on the system-level variables tend to co-vary.

One way of dealing with this problem is to apply the principle of falsification that the most-different systems design is built on. Since the inclusion of many system-level variables makes the regression models unstable, the ambition should be to *exclude* as many system-level variables as possible from the regression models.

This is done by studying interactions between independent and dependent variables *in as varying contexts as possible*. The principle is easy: when an association between the independent and the dependent variable is found in two varying contexts, the more the analytical contexts differ in terms of systemic factors, the higher the number of systemic variables that can be disregarded from the regression model. If, for instance, it can be proven that a high level of education is associated with a permissive attitude towards abortion in Denmark, Russia, Zimbabwe, China, and Egypt we can disregard the system-level variables democracy, secularism, ideological orientation of government in power, and socio-economic development from the analysis since these features vary between the five countries in question.

A similar kind of reasoning can be applied to the use of the most-similar systems design in combination with multilevel modelling. If we only include countries which are similar with regard to a large number of background characteristics, these characteristics do not have to be included in the regression models as control variables, which, in turn, means that the models become more stable. In such cases, of course, the researcher must weigh the utility of building more stable models against having a larger number of cases.⁴

10. CONCLUSION

The most-similar systems design and the most-different systems design are regularly mentioned in educational books in social research methodology in general and in comparative politics in particular. It is not difficult to give examples of studies which, to varying extents, have applied the logic of the most-similar or most-different systems design (e.g. Gerring 2017: 82–97). At the same time it is difficult to find studies where the methods have been applied rigorously, i.e. where the researcher has secured maximal homogeneity or heterogeneity on clearly specified control variables. The main reason for this is that the limited number of countries in the world makes it difficult to apply them in situations where we want to assess the level of association between an independent and a dependent variable. This is particularly the case when we want to construct a most-similar systems design, where the phenomenon of interest varies and all relevant extraneous variables are kept constant. When applied at the system level, a most-different systems design is also sensitive for the problem of too many variables and too few cases, since it is very likely that countries which have the same value on the dependent variable share more than one plausible explanatory characteristic.

In the present chapter, I have argued that the most-similar and the most-different systems designs can be used in many different ways, depending on whether the research question is deductive or inductive and on whether or not the focus is on social phenomena residing at the systemic level or at the sub-systemic level. In line with many other authors, I argue that the most-different systems design can be used both as suggested by Przeworski and Teune (1970), namely for studying variable interactions at the sub-systemic level and comparing the results across systems, but also for research efforts where the dependent variable resides at the systemic level.

Although the methods have been challenged by the introduction of multilevel analysis and QCA, it is easy to predict that researchers will continue to apply the logic of the most-similar and most-different systems designs when conducting cross-country research with a limited number of cases. The methods make use of a variable-oriented approach at the same time as they allow researchers to focus also on the cases, and particularly to account for complex social processes that are difficult to use in statistical analyses. Another advantage with the most-similar and most-different systems designs lies in their ability to eliminate a large number of potentially relevant explanatory variables from further analysis. By carefully matching a small number of cases across a wide range of potential explanatory variables it is possible to exclude a wide range of variables from further analysis.

NOTES

1. The present chapter is based on the following previous contribution: Anckar, C. (2008), On the applicability of the most similar systems design and the most different systems design in comparative research. *International Journal of Social Research Methodology* 11: 389–401. Although many of the arguments raised in that article are still valid, some of the recommendations have been altered.
2. <https://www.freedomhouse.org>.
3. I am indebted to an anonymous reviewer for once pointing this out to me.
4. For an innovative and more detailed discussion of how the most-similar and most-different systems designs can be combined with a multilevel approach, see Denk (2010).

REFERENCES

- Anckar, C. (2004). *Determinants of the Death Penalty: A Comparative Study of the World*. London: Routledge.
- Anckar, C. (2008). On the applicability of the most similar systems design and the most different systems design in comparative research. *International Journal of Social Research Methodology* 11: 389–401.
- Anckar, C. (2014). Why countries choose the death penalty. *The Brown Journal of World Affairs* 21 (1): 7–25.
- Anckar, C. and Fredriksson, C. (2019). Classifying political regimes 1800–2016: a typology and a new dataset. *European Political Science* 18: 84–96.
- Bartolini, S. (1993). On time and comparative research. *Journal of Theoretical Politics* 5: 131–67.
- Berg-Schlosser, D. and De Meur, G. (1994). Conditions of democracy in interwar Europe: a Boolean test of major hypotheses. *Comparative Politics* 3: 253–79.
- Brettschneider, C. (2002). Dignity, citizenship, and capital punishment: the right of life reformulated. *Studies in Law, Politics and Society* 25: 119–32.
- Burt, R. A. (1994). Democracy, equality, and the death penalty. In I. Shapiro (Ed.), *The Rule of Law*. New York: New York University Press, 80–98.
- Collier, R. B. and Collier, D. (1991). *Shaping the Political Arena: Critical Junctures, the Labor Movement and Regime Dynamics in Latin America*. Princeton, NJ: Princeton University Press.
- Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Skaaning, S.-E., Teorell, J., et al. (2018). *V-Dem Codebook v8*. Varieties of Democracy (V-Dem) Project. V-Dem Institute, University of Gothenburg.
- Cronqvist, L. and Berg-Schlosser, D. (2009). Multi-value QCA. In B. Rihoux and C. Ragin (Eds.), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks, CA: Sage, 69–86.
- De Meur, G. and Berg-Schlosser, D. (1994). Comparing political systems: establishing similarities and dissimilarities. *European Journal of Political Research* 26: 193–219.
- De Meur, G. and Berg-Schlosser, D. (1996). Conditions of authoritarianism, fascism, and democracy in interwar Europe: systematic matching and contrasting of cases for ‘small N’ analysis. *Comparative Political Studies* 29: 423–68.
- Della Porta, D. (2008). Comparative analysis: case-oriented versus variable-oriented research. In D. Della Porta and M. Keating (Eds.), *Approaches and Methodologies in the Social Sciences: A Pluralist Perspective*. Cambridge: Cambridge University Press, 198–222.
- Denk, T. (2010). Comparative multilevel analysis: proposal for a methodology. *International Journal of Social Research Methodology* 13: 29–39.
- Dion, D. (1998). Evidence and inference in the comparative case study. *Comparative Politics* 30: 127–45.
- Ebbinghaus, B. (2005). When less is more: selection problems in large-N and small-N cross-national comparisons. *International Sociology* 20: 133–52.
- Faure, A. M. (1994). Some methodological problems in comparative politics. *Journal of Theoretical Politics* 6: 307–22.
- Freundreis, J. P. (1983). Explanation of variation and detection of covariation: the purpose and logic of comparative analysis. *Comparative Political Studies* 16 (2): 255–72. Freedom House. Available at: <http://www.freedomhouse.org>.
- Gerring, J. (2017). *Case Study Research: Principles and Practices* (2nd edition). Cambridge: Cambridge University Press.
- Goertz, G. (2005). Necessary condition hypotheses as deterministic or probabilistic: does it matter? *Qualitative Methods: Newsletter of the American Political Science Association Organized Section on Qualitative Methods* 3: 22–7.
- Goertz, G. and Mahoney, J. (2012). *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.
- Grofman, B. and Schneider, C. Q. (2009). An introduction to crisp set QCA, with a comparison to binary logistic regression. *Political Research Quarterly* 62 (4): 662–72.
- Herron, M. C. and Quinn, K. M. (2016). A careful look at modern case selection methods. *Sociological Methods & Research* 45 (3): 458–92.
- Hug, S. (2013). Qualitative comparative analysis: how inductive use and measurement error lead to problematic inference. *Political Analysis* 21 (2): 252–65.

- King, G., Keohane, R., and Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Knowles, J. B. (2004). Capital punishment in the Commonwealth Caribbean: colonial inheritance, colonial remedy? In P. Hodgkinson and W. Schabas (Eds.), *Capital Punishment: Strategies for Abolition*. Cambridge: Cambridge University Press, 282–308.
- Lieberson, S. (1991). Small N's and big conclusions: an examination of the reasoning in comparative studies based on a small number of cases. *Social Forces* 70: 307–20.
- Lieberson, S. (1994). More on the uneasy case for using Mill-type methods in small-N comparative studies. *Social Forces* 72: 1225–37.
- Lijphart, A. (1971). Comparative politics and the comparative method. *American Political Science Review* 65: 682–93.
- Lijphart, A. (1975). II. The comparable-cases strategy in comparative research. *Comparative Political Studies* 8 (2): 158–177.
- Linz, J. and Stepan, A. (1996). *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe*. Baltimore, MD and London: Johns Hopkins University Press.
- Mahoney, J. and Vanderpoel, R. S. (2015). Set diagrams and qualitative research. *Comparative Political Studies* 48 (1): 65–100.
- McDermott, R. (2002). Experimental methods in political science. *Annual Review of Political Science* 5: 31–61.
- Meckstroth, T. (1975). 'Most different systems' and 'Most similar systems': a study in the logic of comparative inquiry. *Comparative Political Studies* 8: 132–57.
- Mill, J. S. (1970 [1843]). *A System of Logic*. London: Longman.
- Neumayer, E. (2008a). Death penalty abolition and the ratification of the second optional protocol. *The International Journal of Human Rights* 12 (1): 3–21.
- Neumayer, E. (2008b). Death penalty: the political foundations of the global trend towards abolition. *Human Rights Review* 9: 241–68.
- Paine, J. (2016). Set-theoretic comparative methods: less distinctive than claimed. *Comparative Political Studies* 49 (6): 703–41.
- Peters, B. G. (1998). *Comparative Politics: Theory and Methods*. Basingstoke: Macmillan.
- Popper, K. (1959). *The Logic of Scientific Discovery*. New York: Basic Books.
- Przeworski, A. and Teune, H. (1970). *The Logic of Comparative Social Inquiry*. New York: John Wiley.
- Ragin, C. (1987). *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*. Berkeley, CA: University of California Press.
- Ragin, C. (2000). *Fuzzy-Set Social Science*. Chicago, IL: University of Chicago Press.
- Ragin, C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of Chicago Press.
- Rihoux, B. and Ragin, C. (Eds.) (2009). *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks, CA: Sage.
- Rohlfing, I. (2016). Why simulations are appropriate for evaluating qualitative comparative analysis. *Quality and Quantity* 50 (5): 2073–84.
- Sartori, G. (1991). Comparing and miscomparing. *Journal of Theoretical Politics* 3: 243–57.
- Savolainen, J. (1994). The rationality of drawing big conclusions based on small samples: in defense of Mill's methods. *Social Forces* 72: 1217–24.
- Schneider, C. Q. and Wagemann, C. (2012). *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. New York: Cambridge University Press.
- Seawright, J. and Collier, D. (2010). Glossary. In H. Brady and D. Collier (Eds.), *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (2nd edition). Lanham, MD: Rowman & Littlefield, 313–60.
- Seawright, J. and Gerring, J. (2008). Case selection techniques in case study research: a menu of qualitative and quantitative options. *Political Research Quarterly* 61 (2): 294–308.
- Skocpol, T. (1979). *States and Social Revolutions. A Comparative Analysis of France, Russia and, China*. Cambridge: Cambridge University Press.
- Skocpol, T. (1984). Emerging agendas and recurrent strategies in historical sociology. In T. Skocpol (Ed.), *Visions and Methods in Historical Sociology*. Cambridge: Cambridge University Press, 356–91.

- Thiem, A. (2014). Mill's methods, induction, and case sensitivity in qualitative comparative analysis: a comment on Hug (2013). *Qualitative & Multi-Method Research* (Fall): 19–24.
- Thiem, A. and Baumgartner, M. (2016). Back to square one: a reply to Munck, Paine, and Schneider. *Comparative Political Studies* 49 (6): 801–6.

4. Can a case study test a theory? Types and tokens in comparative policy analysis

Keith Dowding¹

1. INTRODUCTION

Since the publication of King, Keohane and Verba (1994), discussion of the methodology of qualitative methods has grown ever more sophisticated. Qualitative methods involve the detailed description and analysis of institutions and events (that I will collectively call ‘items’). No one denies that there is great merit in detailed description and analysis. At the very least, they form the bedrock of social scientific analysis, the basic data upon which, for example, quantitative work rests. The philosophical and methodological issues surrounding qualitative methods concern whether or not they can perform the tasks of testing theories about social phenomena and whether and how good they are at pinning down causal processes. Some qualitative scholars argue that qualitative methods are superior to quantitative methods at pinning down causation or, as they sometimes write, *actual* causation. This chapter analyses one aspect of that debate: in what sense, if at all, can a case study test a theory? In doing so, it touches on, but does not address, the debate on causation. I focus on how a single case study can test theories, and say little about comparative case-study work, though some of the lessons here carry over to the comparative method.

The important lesson of this chapter is that theories are about types. A type is a class composed of token examples (Dowding 2016; Wetzel 2009). A single case study is about a token. A token is a member of a type given some of its characteristics; indeed, any single item can be a member of many types. I say more about types and tokens below. When we discuss causation, we must bear in mind that we might be talking about the cause of a token item in more or less of its detail or about the cause of types in more or less of their detail. So, another lesson of this chapter is that some of the debate over pinning down causation in the social sciences ignores these distinctions. Analysis of types and analysis of tokens belong to separate sets of questions, though our answers to type questions are implicated in our answers to token questions. This chapter is concerned with how questions about token items contribute to our understanding of questions about type.

2. WHAT IS A THEORY?

If we ask whether or not a single case study can test a theory, we must first know what a theory is. We must also be clear what we mean by ‘test’. Tests are tests of theories (we will leave the latter undefined for the moment). However, what is tested when we test a theory is a prediction (or hypothesis) derived from it. This prediction will give us some expectations about how the world should be. Our test concerns whether or not the world is as the theory leads us to expect. For example, theories about party systems derived from Duverger’s original work tell us that

the nature of the party system (the number of parties) is partly dependent upon the electoral system, which includes factors such as the electoral rules (majoritarian versus proportional systems) and the district magnitude (number of seats in each constituency or district) (see Grofman 2006 for a brief review). So, for any given electoral system, the theory has a prediction about the number of parties we should expect to see. Conversely, given the number of parties in any system, the theory tells us what sort of electoral rules we should expect to see. Note, though, that different sets of rules might result in the same number of parties, so the latter calculation will not necessarily lead to a unique prediction.

So, theories, if they are testable at all, make predictions about how the world should be. To be a little more precise, scientific predictions are *conditional*. They say that if condition C holds, then we expect Y; or if conditions C hold, we expect Y with some probability p . Or, if we know that condition Y holds, then we expect that the world is in either condition C or C*. Predictions in the scientific sense are not necessarily about objects or events in the future.² They might be about contemporaneous items or items in the past. A scientific prediction derived from a theory is an inference or implication; it is not a forecast – though it might, with other evidence, be turned into one (Dowding 2017; Dowding and Miller 2019).

Predictions too can also be about types or tokens. A prediction about a token item will give us our expectations about what we expect to see *in that case*. The probability of what we expect to see in that case. But if there are many cases then it will tell us what we expect over each of those cases, and if it is probabilistic the distribution of expected outcomes over those cases. One confusion is that generalizations can be considered theories (as I discuss below) from which predictions can be generated about further cases. Or they can be considered predictions (based on some type of theory – a mechanism) – again I discuss this further below.

Generally speaking, when we ask whether some evidence is a good test of theory, we are asking whether or not that evidence supports or does not support the theory. Evidence that is consistent with a theory may or may not support it. Evidence that is inconsistent with a theory does not support it. In the latter case, we sometimes say, with Karl Popper (1972), that the evidence falsifies the theory. Popper believes you can use evidence to test a theory only in relationship to a rival theory which makes a different prediction. Since he thinks nothing is certain, falsifying theory A means that the evidence reduces the odds that theory A is true relative to theory B. It never means it proves theory A is false (Dowding 2016: 116–27). Evidence that supports theory B relative to A is corroborative, meaning that the odds that B is true are increased relative to theory A. The evidence might have no effect on the odds of A or B being true relative to some other theory C. In fact, C might not have any predictions with regard to that evidence.

Most writers talk about evidence confirming or disconfirming theories. These terms refer to a Bayesian way of thinking about testing theories. When we say that evidence confirms a theory, we are suggesting that it increases our belief that the theory is true. When we say it disconfirms the theory, we are suggesting it reduces the probability that the theory is true.

Some evidence is much stronger than other evidence. Whilst the amount of evidence for or against a theory can often be decisive in whether or not we think a theory is true, the quality of the evidence is often of greater importance. Some evidence is decisive, some only suggestive. Sometimes evidence is consistent with a theory, but does not really affect the probability of it being true, since it is consistent with other, rival, theories. Sometimes the evidence might be so negatively strong that we infer that the theory must be false. One of the arguments for using detailed qualitative case studies rather than large-N quantitative studies is that qualitative evi-

dence is often stronger than that delivered by quantitative studies (George and Bennett 2005; Hall 2006). Large-N studies provide a lot of evidence from cases that have been coded into broad categories, but case studies provide detailed, evidence-specific cases. Sometimes we use cases as examples of a given class of such cases. We examine a token in order to see what we can learn about the type, and how it transfers to other tokens within that type. The claim made about highly detailed studies of such tokens is that this evidence is of higher quality and should be given more weight in our consideration of the theory.

However, some events might be considered to be unique. The ending of the Cold War might seem to be a unique event, with nothing else in history quite like it. The explanation and theory around it might correspondingly also be thought to be unique. For most of this chapter, I will be considering items as though they are specific examples of a given class – token uses to analyse the type. At the end I will consider unique events.

I have not yet stated what a theory is, other than by implying it is something that gives us conditional predictions about how aspects of the world must be. The term ‘theory’ means many things in the social sciences, and indeed even philosophers of science do not use the term consistently, let alone in the same manner as each other (Dowding 2016). I will consider three ways in which the term theory is used (I will state in advance that my preferred version is the third). I ask of each whether a single case study could test it and, if so, how.

The first is theory as invariant generalization. That is, a generalization about the world that always holds (at least under ‘normal’ conditions). As we shall see, these generalizations are invariant because they are identity statements. The second is empirical generalization. Many empirical generalizations in the natural sciences look like invariant ones because they hold almost all the time. But in the social sciences, empirical generalizations often have many exceptions. In both cases, the prediction is that because the generalization holds over all the X cases, it will hold over this particular x case. The third understanding of theory is as a set of propositions that form a complex explanation of why we should expect something to happen. The propositions together form an explanation. I will call this a mechanism. We can note that mechanisms can predict empirical generalizations. As we shall see, invariant generalizations underpin mechanisms.

3. TESTING A THEORY AS AN INVARIANT GENERALIZATION

Philosophers of science often think of natural laws as theories about the world. Boyle’s law states that the volume (V) of a given mass of an ideal gas at a constant temperature is inversely proportional to its pressure (p) that is $pV = \text{constant}$. Boyle’s law is one of a series of gas laws which can be considered to be invariant (Woodward 2000). Invariant generalizations typically *describe* how the world must be. They do not do so by describing any particular state of the world. Boyle’s law does not tell us what the volume, mass or temperature of that amount of oxygen ‘over there’ is – rather, it predicts its volume given its temperature and pressure; or its temperature given the pressure volume; or the pressure upon it given its temperature and volume. The invariant generalization provides a conditional prediction dependent upon measurements of those aspects of the world to which we are applying the law. The invariant generalization is defined relative to a type, and it is given application to token examples of the type.

The gas laws, and many other invariant generalizations, are defined relative to a type that is an ideal. When applied to actual token examples, various other factors need to be considered, and one might then think that, since the claim is that the generalization (the law) is invariant, one case that did not fit the generalization would disprove it. However, one case can only disconfirm – that is, lower the probability. A single case can only disconfirm marginally. That is because Bayesian calculations are based upon our prior expectations and how much we trust new evidence. So, the epistemic version of Bayes's theorem

$$\Pr(A|B) = (\Pr(B|A) \Pr(A)) / (\Pr(B|A) \Pr(A) + \Pr(B|\sim A) \Pr(\sim A))$$

can be given, where $\Pr(B|A)$ is the probability we'd observe result B if A were true, and $\Pr(B|\sim A)$ is the probability we'd observe B if A were not true, with $\Pr(A)$ being our belief that A is true prior to observing B, and $\Pr(\sim A)$ being our belief that A is not true prior to observing B; and so $\Pr(A|B)$ is the posterior probability for A after taking into account the evidence B. What we should understand by all this is that evidence thus affects our degrees of belief – not just how strongly the new evidence will affect our posterior belief but also, crucially, how strong our prior probabilities were. (See Bennett 2009 or Mahoney 2016 for more on Bayes, for example as applied to process-tracing techniques.)³

In order to see why single cases can only affect the probabilities marginally, we need first to think about how invariant generalizations are generally determined. How they are determined fixes our trust in them when we test against other cases. We determine them by measuring parts of the world. We first discover some relationship between the mass, volume and pressure of a certain gas at a given temperature, and realize as we measure the mass, volume and pressure of other gases that a similar relationship exists across different gases. (Usually these are similar and not identical relationships because of measurement errors and other factors that enter into measurement.) We then do some maths, using some simplified assumptions, and work out a mathematical relationship that we deem to be a law or invariant generalization that appears to hold in all the cases that have been studied.

The assumptions and the mathematics we use will also rely upon other theories – laws and invariant generalizations that we have found in other fields. In the case of gas laws, for example, these other factors are derived from kinetic theory, which describes gas as an enormous number of microscopic particles (atoms and molecules) in constant rapid and random motion. These sets of invariant generalizations fit together in coherent systems of equations or descriptions that form more complex theories. These more complex theories are often recognized to be approximations: for example, the gas laws depend on kinetic theory that itself assumes that relativistic effects and quantum-mechanical effects are negligible. If these effects are not negligible, then when we apply Boyle's and other gas laws we need to take account of the non-negligible effects. We then enter into a new set of predictions.

These theories, or what I will call 'models' of the world, are descriptions. Whilst the maths is supposed to be completely descriptive, the narration that helps us grasp what the maths means is usually, if not always, analogical. For example, in kinetic theory we think of microscopic particles as little balls bouncing off each other, but we know in fact they are not quite like that. Nevertheless, in order to grasp our mathematical models of the universe, we often describe them in ways that fit with our perceptual understanding of the macroscopic world. In that sense, models (theories in another sense of the term; see below) are always analogical. This

should not be surprising, because analogy is the most basic form of explanation (Hofstadter and Sander 2013). We usually grasp how X works by thinking it must be a little like how Y works. This helps us to understand how X works when we think we already know how Y works. The explanation of any phenomenon always relies upon the prior understandings of the person to whom the explanation is directed.

Invariant generalizations usually take the form of two sets of propositions either side of an equality sign. This fact tells us that they are equalities, and equalities are generally some form of *identity statement*. Boyle's law is essentially telling us that, given the constant for any specific gas, its mass is pV . The identity does not always constitute the only characteristics of the item on the right-hand side, but it tells us that the item on the left-hand side is constitutive of the item on the right-hand side.

Can we test an invariant generalization with a single case? Well, we first need to note that our evidence for invariant generalizations derives from our discovering empirical generalizations. Empirical generalizations are concerned with phenomenological items. The invariant generalization is theoretical. None of our measurements of a given phenomenon might precisely agree with the invariant generalization, due to measurement and other assumed potential errors. Nevertheless, we have produced a theory that enables us to predict with acceptable accuracy to other cases. Until this one. Because of the problem of measurement error, we would rarely accept that a single case can knock down an invariant generalization. Our previous measures will have formed our prior. Our single case provides contrary evidence; we know there might be mistakes or measurement errors. Hence even if it disconfirms, it will only disconfirm to a marginal extent. We will want to look at other cases before we shift our credence too much. We would want to measure it again, and look at other similar cases. It is true that sometimes crucial experiments can shift our credence by large amounts, but we will still want other confirming evidence.

Furthermore, we will generally want some theoretical reasoning to help us explain the contrary or unexpected findings. Scientists generally work with the rule that you can't beat something with nothing. A contrary finding will remain an anomaly to a theory, rather than a full disconfirmation, until we can think of a better theory that explains both our previous results and the new anomalous finding. One reason we give for anomalous results in the social sciences is that the model does not apply to this case for some reason. It is not that the model is false; rather, it is inapplicable.

Nevertheless, it is hard to shift the view that, logically, a single disconfirming case can knock down an invariant generalization. An old example is that a single black swan can disprove the generalization that 'all swans are white'. In fact, this example is rather misleading, for the reason just given. Finding a black 'swan-like' bird would not disconfirm until we confirmed it is a swan. 'A swan' here must be theorized. Now, of course, the phrase 'all swans are white' is clearly an empirical generalization, not an invariant one. Neither swans nor perceptual colour are theoretical items, so the example is doubly misleading. What could be an invariant generalization along these lines? It cannot be 'all swans are white', since that clearly is not an identity statement. If whiteness is to be taken as constituent of swans, it cannot just be that all the ones we have ever seen are white; we need a set of propositions that together tell us why swans have to be white. It would have to be that, given something that is a defining feature of a swan, and given something that is a defining feature of a white feather, all swans have white feathers.

It is hard to even make a plausible story along those lines, though we can give an example involving animal colouration that does suggest an empirical generalization underpinned by an invariant one. The empirical generalization that no mammal has a striped body and a spotted tail is underpinned by a mathematical model that shows patterns depend on the shape and size of the region where they are formed. Given the developmental stage at which chemicals react to form patterns on the embryo, the maths shows that mammals can have spotted bodies and striped tails, but not vice versa (Dowding 2016: 109). The model cannot be directly tested, but relies upon theory (invariant generalizations) to give an empirical generalization for which there are no extant counterexamples.

Can a single case demonstrate an invariant generalization? If we did find an animal with a striped body and spotted tail, then we would have to examine the theory. But we might start with examining the embryonic formation of that creature rather than query the maths or chemistry of how patterns form in the embryonic stage. That is, we might begin by suggesting not that the invariant generalizations are false, but that, for some reason to be discovered, they do not hold in this case. As stated above, we first investigate why the theory may not *apply* here, before we challenge the theory itself.

Theories are generally formed following empirical findings, and often following findings anomalous to a current theory. Einstein's general theory of relativity was formed not through a series of empirical generalizations that gave rise to a simple invariant one, but by mathematically theorizing how a set of apparently inconsistent empirical results could be made consistent. Furthermore, his work progressed by using thought experiments, not by empirical experimentation. These thought experiments show that if certain things are thus, other things must, logically, be so. Or that, if such things are so, then other things we thought to be the case logically cannot be so. Einstein's theory, then, had a set of claims about the world that were original and had never been measured, one of which was about the degree of curvature of light around heavy bodies. The political methodologists Van Evera (1997: 66–7) and Gerring (2007: 117–18) suggest that Eddington's astronomical observations, which confirmed predictions about the curvature of light around the sun, are an example of a single case that confirmed a theory close to certainty. In fact, however, Eddington's estimations were heavily manipulated and at the time highly controversial; they convinced only those who already accepted general relativity on theoretical grounds (Dowding 2016: 113–14).

Subsequent similar, and other very different, tests strongly confirm Einstein's theory, of course; but a single case study, on its own, can only be confirmatory or disconfirmatory (increase or decrease the probability that a theory is true) and only by the amount of trust and weight of evidence it provides. Broadly speaking, single cases are unlikely to change our views of general theories very much. They might, however, shift our attitude to what we think happens in specific (token) cases enormously. I turn to that below. Here I underline that this fact about shifting our attitude in token cases should not be confused with shifting our attitude in general (type) cases.

4. TESTING A THEORY AS AN EMPIRICAL GENERALIZATION

One might think that testing theory as invariant generalization in political science and public policy studies is irrelevant, since we do not have any such invariant generalizations. I am not sure that is true; what is true is that invariant generalizations in the social sciences are highly

abstract, and either difficult to test or seem so obvious that no one would bother to directly test them. The law of minimum winning coalitions is an invariant generalization. We can state it as saying that in n -person zero-sum games with side-payments, agents will create coalitions that are large enough to win but no larger. Such minimum winning coalitions will consist of a group of agents who (1) agree on a common objective within the game, (2) have sufficient resources to win, (3) are unable to win should any member leave, and (4) divide up the winnings in relation to their resources. Agents thus maximize their gains (Riker 1962: 32–3).

The theory relies upon an understanding of rationality – that is, the preferences of the players conform to certain formal conditions and they have complete information about their own and others' preferences (see Austen-Smith and Banks 1999: ch. 1, for an account of those conditions). As a mathematical model, given its assumptions, it is valid. It is only interesting as political science to the extent that its conclusions are projected on to the world in a manner which we believe can help explain some item – Einstein's equations are similarly projected on to the world to make predictions about the displacement of light around the sun. In the political science case, the theory of minimum winning coalitions gives rise to an empirical generalization about the number of political parties in coalition governments. (Note that the theory as I have specified it above does not mention parties or government; nor does Einstein's maths mention the displacement of light around the sun. Both need application.)

A great deal of work has gone into examining the size principle, which shows that, whilst the empirical generalization holds true more than we would expect from merely random assignment of parties to a government, there are many counterexamples (Laver 1998). In fact, so many counterexamples that some have given rise to other, putatively rival, empirical generalizations. The simple size principle assumes that agents only want to maximize their utility in the form of dividing up some reward specified in the mathematical conditions. In projection to the world of government, this is seen as the share of power, cabinet seats, and so on. But parties also care about policy, and are unlikely to form coalitions with parties that are far apart ideologically. So, the idea of minimum connected winning coalitions (MCWC) came about (Axelrod 1970), the 'connection' referring to being close in ideological space. Of course, 'ideological space' is a theoretical construct, which again has to be projected on to the world and measured (Laver and Hunt 1992: ch. 1). Constructing measures of ideology is normally conducted through measuring positions over a series of issues. These issues are then formed into a single ideological dimension. But we can also turn them into two- (or more) dimensional ideological space. Or just think about n -dimensional issue space. This gives rise, in projection on to the world, to new predictions (Laver and Shepsle 1996).

We can see, therefore, that, as theory gets more complex, testing it becomes more problematic. Each formal theory is valid in its own terms, but in order to empirically test it, we need to project it on to the world. That projection will give rise to empirical generalizations. Testing these empirical generalizations requires us to see how often the world resembles the claim of the generalization. If the generalization does better than random, then we can say it is confirmed. However, if it does only a little better, we might be sceptical that the theory, the model that generated the invariant generalization which was then projected, tells us much about the world. What we can do, however, is improve the theory by adding more elements to it – that is, make it more complex – to give predictions that better project on to the world. We must note, however, that empirical generalizations can only be confirmed – can only be shown to hold to the extent that we find the theory underlying the generalization helps us understand the world – by looking at numerous cases. Testing theory in terms of empirical generalization

can only be done by large-N analysis. Single case studies are of only marginal interest in this form of testing.

Nevertheless, we have to be wary in claiming that adding more elements to the theory, making it more complex, will generate better-fitting projections. We can find projections that virtually always seem to be the case, but these would often be trivial and uninteresting. ‘All democratic governments are composed of a single-party or a coalition of parties’ might have few counterexamples, but is trivial. To the extent it is an invariant generalization, it is true because of the meaning of the terms that compose it – ‘democratic governments’, ‘party’, ‘coalition’. All invariant generalizations are true because they categorize identities, but the interesting ones use abstract theoretical terms that fit into broader understandings, not ones that merely identify the macro phenomenon we are trying to explain.

‘Party government is always composed of a party or a set of parties’ is true by definition. Being composed of a party is part of the definition of ‘party government’. We noted above that invariant laws in physics are (often at least) identity statements, but they are interesting because, prior to discovery, the identity was not obvious, and furthermore that identity is partly made up of theoretical, not just observational, items and also fits with a body of other such invariant generalizations. That can be true of mathematical models in the social sciences, but their projections on to the world and the empirical generalization generated from them are less obvious and secure. What we learn is that we are only interested in generalizations that are underpinned by a theory that we consider to possess merit, in that it can explain why we see the patterns we do in macro phenomena and can project at least reasonably well.

What do we mean ‘underpinned by a theory’? In this section we have been considering whether a single case can test a theory where ‘theory’ means ‘empirical generalization’. We have found it cannot. But we are only interested in such empirical generalizations that are ‘underpinned by theory’. ‘Theory’ here must mean something other than empirical generalization. It might mean invariant generalization as we have seen, but equally it might mean something else. It might mean, as we have been describing it, a model, perhaps a mathematical one. It is these models that generate the generalization as predicted. These models can be thought of as descriptions or ‘mechanisms’ that operate in the world. These models, and not just the predictions they produce, can themselves be projected on to the world. They would then describe mechanisms. Can case studies test these mechanisms? I turn now to this question.

5. TESTING A THEORY AS A MECHANISM

In political science, qualitative methodologists, largely, though not entirely, in response to King et al. (1994) have defended the ability of qualitative methods to test (and generate) theories. The qualitative methodology that is defended in this regard usually goes under the name of process tracing (Beach and Pedersen 2016; Bennett and Checkel 2015; Collier 2011; George 1979; George and Bennett 2005). The defence rests upon the claim that theories in the social sciences should be seen as mechanisms rather than generalizations. Theory can mean generalization and can mean a mechanism; and it can mean either in both the social and natural sciences (Dowding 2016). I also claim that in both, invariant generalizations underlie mechanistic explanation (see also Waldner 2012) and mechanisms generate empirical generalizations. As we have seen above, single cases cannot effectively test either invariant or

empirical generalizations. However, the claim of process tracing is not (usually) about testing generalizations as such, but about testing mechanisms.

Moreover, the claim of process tracers is that mechanisms provide more robust explanations than generalizations do. Indeed, we might claim that empirical generalizations are not really explanatory at all. In what sense does saying ‘that animal’ is a leopard explain why it has spots on its body and a striped tail? Being told that virtually all leopards have spots on their bodies and striped tails only helps us to identify that animal as a leopard. No more. To explain why leopards have spotted bodies and striped tails, we need to supply some kind of evolutionary story about the advantage to that kind of predator of such colouration. Then we might tell the story of the pigmentation about why spots and stripes tend to go together. Each story is a mechanism. The first is an evolutionary story, the second a chemical developmental one. Now we note that the evolutionary explanation of why having spots or stripes advantages predators does not tell us why leopards (as a type) have the sort of pigmentation they do, let alone why ‘that leopard’ (the token) has the precise pigmentation it does. That requires the chemical story, and the precise development of the foetus for that particular leopard. The evolutionary explanation gives us an explanation of pigmentation at one level of analysis, the precise pigmentation at another. And they answer subtly *different* questions. They also utilize different mechanisms. I will argue that large-N and case studies in the social sciences likewise answer subtly different questions, but might both involve similar mechanisms.

Each are stories about how certain forces tend to lead to the empirical generalizations we witness. In the coalition model mentioned above, the generation of the prediction about MWC or MCWC was a story about how certain forces (utility of winning, possibility of side constraints, ideological location) cause certain outcomes to be expected. And those outcomes are, to use statistical terms, the dependent variables: the outcomes to be explained by the mechanism. Invariant generalizations are relationships that do not vary, and underpin our basic understanding of our models of the mechanisms. They might be worked out by seeing empirical generalizations, and they might themselves be underpinned by more basic mechanisms (kinetic theory underpinning the assumptions of the gas laws; utility theory underpinning the assumptions of coalition theory), but they are descriptions nonetheless. They are not themselves causes, though they enter into causal stories in a major way.⁴

Mechanistic stories are causal stories. The mechanism explains how it comes about (why) leopards have spots and stripes, why it is that coalitions tend to be minimum winning or minimum connected winning. They provide the structure, or the *structural causal* story. Structural causal stories are about types. ‘Leopard’ is a type of animal. Each actual leopard (and each counterfactual leopard) is a token example of the type. Mechanisms are explanatory of types. They are explanatory of any given token to the extent that the mechanism applies to that token. And the extent to which it applies to any given token can vary. Process tracing is good at examining how far any given mechanism applies to token examples. Through that application, it can help explain how useful the type category is.

Before I elaborate further on why process tracing is good at these tasks, I need to say a little more about what a mechanism is. The intuitive idea of a mechanism and how it helps to explain outcomes is fairly clear. Indeed, when we think about mechanisms, we probably think of simple ones such as the working of a clock or the cogs that drive a wheel. (In fact, we can note, the very idea of a mechanism as a form of social explanation is an analogy to physical mechanisms like a lever or clockwork that we think we already understand.) I will use the example of a lever to explain the basic idea of a mechanism.

A lever is a rigid beam that pivots on a fulcrum. Ignoring friction, and assuming the beam is perfectly rigid, the power into the lever equals the power out. The ratio of output to input force is given by ratio of the distance from the fulcrum to points of force. The ratio is an invariant generalization (the law of the lever). A lever can enable us to shift weight that otherwise we could not. The advantage it gives is determined by the torque (T) at the fulcrum. We vary the distance of the force from the fulcrum. $T_1 = F_1 d = T_2 = F_2 d^*$ where F_1 is the input force, F_2 is the output force and d and d^* are the perpendicular distances between the fulcrum and each force. As the torque must be balanced $T_1 = T_2$, so $F_1 d = F_2 d^*$ and the lever gives a mechanical advantage (M) of $M = F_2 / F_1 = d / d^*$.

The mechanism includes the motive force that is applied, the set-up as described and the invariant generalizations given in the mathematical formula. Invariant generalizations underpin the mechanism.⁵ The type causal story is the general set-up as described. The causal story given in any token example is the actual distances and actual forces. So, we can see in a mechanistic explanation that we have a token causal story (the story of any actual example), and a type with those precise features, whilst the mechanism itself is also a type. We can note that the token of any given example will ignore some features of the actual example. In the set-up of the lever here, I have assumed perfect rigidity and no friction. Actual token examples will include some give in the lever and some friction. But as long as those features are not too large, the mechanism as described in theory can explain the outcome. If we want a finer description and an explanation of the actual outcome, we will need to include those features in our explanation. It all depends on how finely grained a description we need for our predictions and explanations.

The simple example I have given of a mechanism is designed to give us an idea of what a mechanism is, given the wealth of incompatible definitions that exist in the literature (see Hedstrom and Ylikoski 2010 or Beach and Pedersen 2016 for a review). Woodward (2002) offers a more formal account of how a model describes a mechanism, summarized by Hedstrom and Ylikoski (2010: 51, table 1):

A model of a mechanism (a) describes an organized or structured set of parts or components, where (b) the behavior of each component is described by a generalization that is invariant under interventions, and where (c) the generalizations governing each component are also independently changeable, and where (d) the representation allows us to see how, by virtue of (a), (b), and (c), the overall output of the mechanism will vary under manipulation of the input to each component and changes in the components themselves.

As Woodward acknowledges, the counterfactual effects of the manipulation are underpinned by invariant generalizations – as we see in the lever example. But an important aspect of a mechanism is that it is composed of entities and activities – the latter being the things that entities do (Craver 2006: 371). Activities are understood at least partly in terms of the manipulability of the variables. We can manipulate the value of one variable in the description of the mechanism by manipulating another (Pearl 2009; Woodward 2003).

Also important is how mechanisms are organized: that is, how the activities are arranged spatially, temporally, and hierarchically. In the social sciences, because of the ‘activity’ idea, we see human agency as a key component. It is for that reason that some writers want to reduce mechanistic explanation in the social sciences to mechanisms about human psychology (Elster 2015). However, in political and policy science, general theories (mechanisms) are *institutional* – how rules affect the incentives for behaviour – so it is these rules that govern the

activity that form the mechanism. So, the precise rules of coalition formation – for example, the role of the *formateur* – can play a key part in the formation of coalition government across different systems (Strøm and Nyblade 2009). It is not that there is no role for political psychology, but standardly in politics and policy analysis important explanatory mechanisms are institutions. We tend to hold psychological factors as fixed – in the form of stable distributions across the population of actors – to examine the effects of varying institutions upon outcomes.

There are two ways we can think about ‘testing a mechanism’. The first is to see whether the purported mechanism works as it is supposed to do. Do levers actually operate as intended? With the simple lever example, when we apply the mathematics from the ideal case, as when we apply the mathematics of the ideal gas case, we need to consider local parameters such as the actual rigidity of the lever, frictional effects, and so on. Given local parameters, each time we measure whether the lever’s effects are what we expect, we provide a test.

The second way in which we might test a mechanism is to see whether the purported mechanism is what explains the outcome in *this token case*. We are not testing the mechanism as a type; we are seeing if the mechanism is applicable to *this* token case. With the lever example, we might ask how a given weight was lifted up on to a rock shelf. Was it moved by a lever or was it raised by some other process (say, lifted up by a flood). The first way of testing a mechanism is testing it as a *type*. Does the type ‘lever’ have the form it is supposed to? The second way of testing is to see if the mechanism explains the token case. Here we are not really testing a theory (in the form of a mechanism) in the manner in which theory-testing is usually understood in science: that is, a type. Rather, we are seeing if the theory applies to this particular case, the token. The test is not of the mechanism itself (is the mechanism true?), but whether it applies to this given case.

Many studies on the advantages and disadvantages of large-N quantitative and small-N or case-study qualitative seem to talk past each other, since they do not acknowledge these two different sorts of tests. Large-N analysis typically correlates a set of independent variables with the dependent variable then theorizes the mechanism that relates them. Manipulation of some of these independent variables would then give further tests that the mechanism works as it has been theorized. Typically, in single or small-N case studies the test is rather whether or not the conditions specified in the mechanism hold in the token case. In both large- and small-N given what we find we might further theorize about the nature of the mechanism – either in the individual case, or across all cases. Indeed, empirically testing theories always involves further theorizing.

We should not confuse these two sorts of tests of theories. Perhaps confusion over theory-testing with case studies arises because the case is often considered to be unique. The ‘theory’ that is being tested *is the theory* of this unique case. It might seem, therefore, that for unique cases, testing whether a theory is applicable to the case *is* testing the theory. However, that is not so. Instead, it confuses testing a token with testing a type. Theories always apply to types. The unique case is a token of a type of which it is the only actual member. It is not the only member, however, for there are counterfactual members. It is thinking about the counterfactual members that enables us to work out which aspects (variables) of the story are theoretically important. Importance here is defined in terms of the probability that the outcome would be of a different *type* if the variables were manipulated in some manner.

6. UNIQUE CASES

Historical process tracing, then, is at the token level and provides proximate explanation of a given case. No single case study can test a mechanism in the sense of showing that the mechanism does not work. All it can ever do is to suggest that the mechanism either appears to operate in the situation being studied or does not appear to operate in the situation that is being studied. Process tracing uses its four ‘tests’ for evaluating inferences, which are made in the context of evaluating whether a larger mechanism operates in a given case study. Most studies compare the evidence for rival mechanisms. Collier (2011) neatly illustrates the four tests with the Silver Blaise Sherlock Holmes story, and process tracing is often compared to detective work or forensic science. However, as Mahoney (2012) acknowledges, in social scientific practice, it is much harder to be sure how evidence matches up to the four tests. These tests only show how relevant given mechanisms are to the case under study.

The claim I am making about the type–token distinction being a key aspect of the nature of the explanation that is being offered can be obscured by the fact that qualitative case studies are of ‘unique’ events. If they are unique, then the explanation of the specific token has also to be unique. However, that does not mean that type explanation is not what we require. If someone asks ‘how on earth did you get that weight up on that high shelf?’, we can tell the story in all its detail of how that was achieved. But if the question is about how someone as weak as you managed it, then the answer required might be that of the ‘principle of the lever’. Unique events are not explanatorily special in any way if the term ‘unique’ simply means ‘individuated’. In that sense, all events are individuated by space–time coordinates at the very least (Fetzer 1975; Tucker 1998: 62) and there is no reason why we cannot have scientific explanation of individuated events.

We tend to think of historical events involving humans as unique. They are unique in the trivial sense that they only happen once in human history. In that sense they are individuated. But what is usually meant by unique is that there is no other event like it in human history. That does not change the fact that the unique event is not the only member of its type, it is just that it is harder for us to consider the counterfactual questions to work out what is important in this event. It is sometimes argued that we cannot apply scientific reasoning to these events, nor generalize from them. One way of thinking about this is to consider the complexity of such unique events. We might be able to say something about them constrained by what we know from other events. We can constrain the action of the individuals within them, but not predict the actual outcomes – or rather not the actual outcomes at some level of granularity. If we want a really detailed story of every aspect of an outcome, then it might well apply only to that unique event. However, that explanation will still involve type-level analyses of it, especially of any aspect that is a purported mechanism.

Perhaps we can acknowledge that some events are so complex, there are so many interactions between the variables, that we cannot model the mechanism that leads to the outcome. The complexity means that if the event were to unfold over and over again, the same outcome would not always occur. It is that sort of complexity that leads researchers to computer simulation techniques, such as agent-based modelling, where the modellers themselves sometimes do not fully understand the process or mechanism that leads to outcomes.⁶ Some of these interactions, modelled at the limits of appropriate granularity, are probabilistic, such that the same outcome will not occur if the model was played time and again. At best, we can give some probability distributions over given outcomes.

But this shows us that we can scientifically study unique events. To be sure, our scientific study would involve modelling the individuated event with all of its important features, and running the event over and over again to see the likelihood of a particular outcome. Say it comes about only 1/10,000 runs. All we can say is that this outcome is highly unlikely. We might not be able to specify how it came about; we do not really know what the full set of interactions were that led to it. It is 'unique' in that it is individuated, unlikely and not fully understood. But we have still studied it scientifically. We can still put it under an empirical generalization (under these initial conditions it would occur one in 10,000 times). We might not understand the process, but we can say that it is a possible outcome given a mechanism that seems to lead to a different outcome most of the time.

For testing the mechanism that leads to the type of outcomes of the unique event 'end of the Cold War', however, we would be most interested in the aspects of the actual situation which, when present in the simulated version, most often lead to that type of outcome. These are the most important aspects. One confusion in debates over applying social science to history is that historians are most interested in the items which very rarely lead to a particular outcome, whilst social scientists are interested in the items which would often tend to lead to that particular outcome. This confusion is compounded by the fact that historians tend to be interested in detail – high granularity of the token – whilst social scientists tend to be interested in the general form – low granularity of the type. Recognizing that these two forms of analysis are both important, and asking different questions, can avoid some of the disputation.

Our lack of understanding of uniqueness is epistemological. We cannot really claim that, ontologically, there is something different about unique events that mean they cannot be studied scientifically, only that they are so complex we cannot really grasp or understand the process by which they unfold. In fact, of course, if the real-world outcome occurred so infrequently in the agent-based model of it, we would almost certainly think we had mis-specified some initial conditions and look for more plausible assumptions to put in our algorithm. We would try to run models where the 'correct' outcome appeared more often and then our understanding of that model process would be projected on to the world.

Detailed historical case studies, especially in the process-tracing form, see whether some general mechanism works in the case, but also detail at high granularity the specific events that led to the precise outcome. Some of these events, or variables, might have been highly unlikely. When we study historical events, we try to pick out the features that led to the outcome we see. We try to understand how the outcome 'had' to have come about. We look for the necessary and sufficient conditions for the outcome. The danger is that we might end up telling 'just so stories' (Evangelista 2015: 155). But we have to remember that Stephen Jay Gould's criticism of 'just so stories' was applied to evolutionary thinking. Gould's criticism is that an evolutionary biologist finds some feature of a creature and has to account for it as being either to its evolutionary advantage or some holdover from some previous fitness advantage. Gould (Gould 1997; Gould and Lewontin 1979) suggests that there might be 'spandrels' that never served any purpose, but never detracted from the animal's fitness enough to be lost.

A 'just so story' is simply another name for a hypothesis. The problem with a unique event, such as 'the end of the Cold War', is that the hypothesis for why each of its particular contributing events are necessary and jointly sufficient for the outcome cannot be directly tested, for we cannot run the unique event again. The four process-tracing tests, along with Waldner's (2015) 'completeness standard', are there to see which mechanism best fits the narrative.

When we narrate the events of a given set of historical circumstances, we cannot help but look for the key actions, events, happenstances, and so on, that we feel determined the outcome. But the determination is from hindsight. When we look at what causes some outcome, we look for some combination of necessary and sufficient conditions. If we identify the conditions that were necessary and sufficient for the outcome, we have shown why it happened. And, given those conditions, it ‘must have happened’ (they are necessary and sufficient). If those events had not occurred, then the observed outcome would not have followed. However, saying these events determined the outcome does not mean that we do not think a different outcome could not have occurred. Often in historical narratives some of the conditions are highly contingent or highly unlikely.

Think back to my attempt to ontologically characterize uniqueness. We might think that our set of necessary and sufficient conditions is highly unlikely and so the outcome was highly unlikely. Or perhaps it is more likely that we think that only one or two happenstances and their interaction were highly unlikely, but that they were key to this particular outcome. And we note ‘this particular outcome’ is specified at a given level of granularity. At a lower granularity, such unique events can often be fitted within a broader empirical generalization. Britain’s declaration of war on Finland in 1941 is an oft-cited counterexample to the empirical generalization that democracies do not go to war with each other; but it can be easily explained by the Finns’ cooperation with the German attack on the Soviet Union, in order to regain territory previously lost to the Soviet Union, factors which can be fitted into other empirical generalizations and processes.

Where we have to be careful in looking back is when the necessary and sufficient conditions involve the intentions of actors. People tend to narrate their lives and retrospectively give themselves intentions that they might not have recognized at the time. Actors might say they acted thus in order to achieve the outcome which transpired, whereas in reality they were simply muddling through. It is also the case, though, that our intentions might not always be conscious and we might choose – not through conscious rational decision making, but not entirely by chance either – certain courses of action that lead to favourable outcomes. In these cases, reading back intentions is not entirely fanciful.

The ‘just so’ pitfall can cause us to read into a specific token event some general mechanism to explain the outcome, where that mechanism is irrelevant. The cause of the outcome is happenstance or a by-product of the process that would not occur if the event were to be replayed many times. The ‘just so’ problem is the opposite of the specification problem. The specification problem is that we can tell several stories that would equally lead to the same outcome. So, we have several structural-type explanations for which the evidence is equally applicable. The ‘just so’ problem is that we do not need a structural story at all, or, rather, the only relevant story is the proximate token one. It is at a level of granularity beyond the structural-type explanation of that token specified as a type.

7. CONCLUSIONS

King et al. (1994) argued that there is one logic of inference limiting the scope of explanation that qualitative work could achieve. Many scholars have subsequently examined the worth of case studies and how they can contribute to our knowledge. Some have argued that case studies can provide stronger causal inferences than large-N quantitative work. As part of that

argument, some have suggested that quantitative analysis relies upon seeing explanation in terms of generalizations or laws, whilst qualitative work has a superior account of causation in terms of mechanisms.

I have argued that invariant generalizations underlie all mechanisms and that mechanisms ought to imply empirical generalizations. I have examined how case studies can test theories understood as either invariant or empirical generalizations, or as mechanisms. We find that understood as a test of whether the theory is true, case studies can have only marginal impact relative to large-N for theory understood as some form of generalization. Nor is a case study likely to demonstrate that a mechanism cannot be true, unless it unpacks a particular theoretical claim about how one variable affects another. Even then, in the social sciences, a single case study can only throw doubt upon the claim; replication and other studies would provide further evidence. What case studies can do, however, is show that a purported mechanism does not hold in a particular token case. By comparing case studies, we might demonstrate that though they have the same type of outcome, that outcome can be reached by different mechanisms.

When we study a unique event, such as the ending of the Cold War, the very idea of testing a theory is to ask which of several purported mechanisms was actually the one which operated in that case. We are interested in the actual causal mechanism, not the most likely causal mechanism. We might well find, if we were to rerun history again and again, that the actual cause of the end of the Cold War is one of the least likely possibilities.

In science we are normally interested in those causes of items that are most likely. We are interested in explanation of type, of structural or ultimate causation. However, when we examine our actual history, we are interested in the explanation of the actual token, not the most likely explanation of items of that type. Hence debate about the worth of case studies testing theories has been problematic, with writers at times talking past each other. We must keep in mind the type–token distinction in social analysis, and be aware, when we make claims about items, whether we are taking the item as a type or as a token. Generally, case studies are not the best way of testing theories about types, but they are the best means to discover whether or not purported explanations of type items apply to a particular token case.

NOTES

1. I would like to thank Anne Gelling, Darren Lim, Charlie Miller, and Marija Taflaga for their comments on an earlier version of this chapter.
2. Theories might be about events – such as the mortality rate under certain conditions – or about objects – such as the number of political parties in a given democracy. The difference between the two with regard to scientific predictions sometimes matters, but in this chapter I will use the term ‘item’ to refer both to an event or an object where the difference between the two does not matter.
3. In the social sciences, the claim is that detailed studies can shift our views more than marginally. We ignore that claim here, since it is not usually applied to invariant generalizations, but to mechanisms. Even there, we need to be careful when making this claim.
4. They are necessary conditions for the causal story.
5. As explained in Dowding (2016), invariant generalizations underpin mechanisms; mechanisms give rise to empirical generalizations. Some empirical generalizations (especially in the natural sciences) seem so invariant that they are confused with those generalizations that underpin the mechanism that generates them. In these cases, we sometimes ignore the mechanism. ‘Metals heat when expanded’ would be an example of such an empirical generalization.
6. ‘Not fully understand’ might simply mean that they cannot mathematically model it at the granularity required to produce the prediction.

REFERENCES

- Austen-Smith, D. and Banks, J. S. (1999). *Positive Political Theory I: Collective Preference*. Ann Arbor, MI: University of Michigan Press.
- Axelrod, R. (1970). *Conflict of Interest*. New York: Columbia University Press.
- Beach, D. and Pedersen, R. B. (2016). *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. Ann Arbor, MI: University of Michigan Press.
- Bennett, A. (2009). Process tracing: a Bayesian perspective. In J. M. Box-Steffensmeier, H. E. Brady, and D. Collier (Eds.), *The Oxford Handbook of Political Methodology*. Oxford: Oxford University Press, 1–23.
- Bennett, A. and Checkel, J. T. (Eds.) (2015). *Process Tracing: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press.
- Collier, D. (2011). Understanding process tracing. *PS: Political Science and Politics* 44: 823–30.
- Craver, C. F. (2006). When mechanistic models explain. *Synthese* 154: 355–76.
- Dowding, K. (2016). *The Philosophy and Methods of Political Science*. London: Palgrave Macmillan.
- Dowding, K. (2017). So much to say: response to commentators. *Political Studies Review* 15: 217–30.
- Dowding, K. and Miller, C. (2019). On prediction in political science. *European Journal of Political Research* 58 (3): 1001–18.
- Elster, J. (2015). *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences* (revised edition). Cambridge: Cambridge University Press.
- Evangelista, M. (2015). Explaining the Cold War's end: process tracing all the way down? In A. Bennett and J. T. Checkel (Eds.), *Process Tracing: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press, 153–85.
- Fetzer, J. H. (1975). On the historical explanation of unique events. *Theory and Decision* 6 (1): 87–97.
- George, A. L. (1979). Case studies and theory development: the method of structured, focused comparison. In P. G. Lauren (Ed.), *Diplomacy: New Approaches in History, Theory and Policy*. New York: Free Press, 43–68.
- George, A. L. and Bennett, A. (2005). *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- Gerring, J. (2007). *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.
- Gould, S. J. (1997). The exaptive excellence of spandrels as a term and a prototype. *Proceedings of the National Academy of Sciences USA* 94: 10750–10755.
- Gould, S. J. and Lewontin, R. C. (1979). The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London* 205 (1161): 581–98.
- Grofman, B. (2006). The impact of electoral laws on political parties. In B. R. Weingast and D. A. Wittman (Eds.), *The Oxford Handbook of Political Institutions*. Oxford: Oxford University Press, 102–18.
- Hall, P. A. (2006). Systematic process analysis: when and how to use it. *European Management Review* 3 (1): 24–31.
- Hedstrom, P. and Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology* 36: 49–67.
- Hofstadter, D. and Sander, E. (2013). *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. New York: Basic Books.
- King, G., Keohane, R. O., and Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Laver, M. (1998). Models of government formation. *Annual Review of Political Science* 1: 1–25.
- Laver, M. and Hunt, W. B. (1992). *Policy and Party Competition*. London: Routledge.
- Laver, M. and Shepsle, K. A. (1996). *Making and Breaking Governments: Cabinets and Legislatures in Parliamentary Democracies*. Cambridge: Cambridge University Press.
- Mahoney, J. (2012). The logic of process tracing tests in the social sciences. *Sociological Methods and Research* 41 (4): 570–597.
- Mahoney, J. (2016). Mechanisms, Bayesianism, and process tracing. *New Political Economy* 21 (5): 493–9.

- Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd edition). Cambridge: Cambridge University Press.
- Popper, K. R. (1972). *The Logic of Scientific Discovery* (6th impression (revised)). London: Hutchinson.
- Riker, W. H. (1962). *The Theory of Political Coalitions*. New Haven, CT: Yale University Press.
- Strøm, K. W. and Nyblade, B. (2009). Coalition theory and government formation. In C. Boix and S. C. Stokes (Eds.), *The Oxford Handbook of Comparative Politics*. Oxford: Oxford University Press, 1–27.
- Tucker, A. (1998). Unique events: the underdetermination of explanation. *Erkenntnis* 48: 59–80.
- Van Evera, S. (1997). *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press.
- Waldner, D. (2012). Process tracing and causal mechanisms. In H. Kincaid (Ed.), *The Oxford Handbook of Philosophy of Social Science*. Oxford: Oxford University Press, 65–84.
- Waldner, D. (2015). What makes process tracing good? Causal mechanisms, causal inference, and the completeness standard in comparative politics. In A. Bennett and J. T. Checkel (Eds.), *Process Tracing: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press, 126–52.
- Wetzel, L. (2009). *Types and Tokens: On Abstract Objects*. Cambridge: Cambridge University Press.
- Woodward, J. (2000). Explanation and invariance in the special sciences. *British Journal for the Philosophy of Science* 51: 197–254.
- Woodward, J. (2002). What is a mechanism? A counterfactual approach. *Philosophy of Science* 69 (S3): S366–S377.
- Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

PART II

THEORETICAL CHALLENGES

5. Comparing policy processes: insights and lessons from the Advocacy Coalition Framework research program

Daniel Nohrstedt, Christopher M. Weible, Karin Ingold and Adam D. Henry

1. INTRODUCTION

Scholars strive to understand why public policies differ across contexts, how public policy outputs vary across different institutions, why and how policy processes differ, and under what conditions policies may improve the well-being of societies. Addressing these questions inherently requires a comparative perspective, as well as collective efforts by scholars working across different substantive cases and contexts. Over time several theoretical frameworks epitomize such efforts. This chapter introduces one of these frameworks: the Advocacy Coalition Framework (ACF).

This chapter describes the ACF as a research program for supporting comparative public policy research (Jenkins-Smith et al. 2017). First, we discuss the key concepts and assumptions of the ACF and how they support comparative empirical research to advance our understanding of policy processes around the world. Second, we provide a summary of comparative research conducted under the ACF and identify the insights and lessons from these efforts for informing and advancing methodology for comparative policy analysis.

The ACF was originally developed to fill the need for a framework that could overcome the limitations of the stages heuristic, or policy cycle approach. Building from the seminal work of Sabatier and Jenkins-Smith in the 1980s, the ACF has evolved into an established international research program with several hundred applications worldwide (Jenkins-Smith et al. 2017; Sabatier 1986, 1988; Sabatier and Jenkins-Smith 1993). The ACF helps researchers explain some of the most foundational aspects of the policy process, including how and why policy actors collaborate through “advocacy coalitions” to influence and affect public policy and under what conditions policy-oriented learning and policy changes are more or less likely to occur.

Comparative work involving a large number of scholars from around the world has enabled advances on all these fronts. For instance, today we know considerably more about why and how advocacy coalitions form and maintain themselves than we did just two or three decades ago. These insights result from the combined efforts of scholars developing and applying novel methodological approaches – from process tracing to statistical network analysis – in formulating, testing, and advancing theoretical arguments.

Following Ostrom (2005) and Laudan (1978), the essential component of a research program is a set of concepts and assumptions regarding the relationships among phenomena, which provides a common vocabulary for promoting continuous exchange and collaboration among scholars working in different cases, contexts, and political-administrative systems. The

ACF thus provides a platform for *implicit* comparative policy process research as it assists and encourages scholars in constructive criticism, collaboration, and exchange – all around the same set of concepts, hypotheses, and methodological practices. Such research efforts help maintain an enduring dialogue for advancing theoretical development. In this regard, the status of the ACF as a research program roughly follows the logic of method of structured and focused comparison, whereby studies (i) address similar research questions to promote comparison, and (ii) focus on certain aspects of a class or category of cases (George and Bennett 2004: 67). Studies are then indirectly compared and contrasted through ongoing scholarly interactions – via publications, conferences, teaching, and so on – which maintain critical dialogue, exchange, and collaboration among researchers.

While the ACF research program provides a platform for ongoing implicit comparison, it has also generated substantial *explicit* comparative research through within-case and across-case comparisons. In aggregate, these contributions have been an important empirical base for introducing and exploring new concepts, developing and testing novel methodologies, and empirically validating the basic assumptions and propositions of the framework. In this perspective, the ACF has supported a considerable amount of comparative policy process research, paving the way for new and exciting empirical insights across a range of cases, policy issues, and political systems, as well as novel theoretical insights about policy subsystems, coalitions, and processes of learning and policy change.

These advances aside, the evolution of the ACF research program has also encountered significant challenges and more work is needed to exploit the ACF's full potential as a comparative research program. For example, as we discuss in this chapter, clear definitions of concepts do not always lead to consistent and comparable operationalizations. Measures and ways to identify the most important contextual attributes for describing and explaining variation in coalitions, learning, and policy change remain elusive. Yet, a message of this chapter is also one of optimism for the ACF's research program to offer a viable approach for comparative policy process research and for contributing to the global reservoir of knowledge about the policy process.

2. ELEMENTS OF COMPARISON WHEN APPLYING THE ACF

Following the ACF, coalitions form, and policy learning and change happen within policy subsystems. Under the guidance of the ACF, policy subsystems are the primary units of analysis, and are characterized by one substantive issue (e.g. health, water quality, public transport), by a geographical scope (most often related to the area of a political jurisdiction at the local, regional, or national level), and by actors that interact regularly and seek influence within these issues and geographies. These actors form advocacy coalitions and compete over the framing, design, and implementation of public policies. Finally, how actors behave and coalitions form and evolve has an impact on how policy learning and policy change develops for a subsystem. For an illustration of the policy process as defined by the ACF, see the flow diagram in the Appendix to this chapter.

Deduced from the ACF's basic assumptions and related to the three core components of the ACF (coalitions, policy change, and policy learning; see Jenkins-Smith et al. 2017), we can identify the elements that can be and typically are compared when conducting ACF research.

To gain insights about formation of coalitions, learning, and change, the ACF focuses research on actors, i.e. individuals and organizations interested in public policy making. Actors are agents of change in subsystems and are guided by their goals. Due to cognitive limitations, policy actors tend to simplify the world based on their belief systems. These belief systems have a three-tiered structure with deep core beliefs (normative values), policy core beliefs (normative and empirical beliefs bound to the policy subsystem), and secondary aspects (instrumental means for achieving desired outcomes within the subsystem). ACF research typically emphasizes *actors' beliefs, which notably provide their goals in the subsystem* (Heikkilä et al. 2019; Weible et al. 2019). However, it is often unclear how these goals can be achieved; hence, the ACF also emphasizes the deployable resources and executable strategies by actors to influence policy subsystem issues.

Individuals, their beliefs and actions, provide important information for the identification of coalitions (Henry 2011; Ingold 2011). When interested in comparing different coalitions, one major element of comparison is the degree of coordination among actors (Matti and Sandström 2011; Weible and Sabatier 2005). Based on similar beliefs, members of advocacy coalitions tend to coordinate actions to increase their individual, but also their joint impact on policy decisions. *Coordination among policy actors* is one key element of ACF research and creates a complex network of interactions among individuals and organizations (Weible and Sabatier 2005).

Moving on from individuals and their interactions within coalitions, the ACF highlights interactions *between* advocacy coalitions. The ACF assumes that public policy is shaped by competition, conflict, or negotiation among one or more advocacy coalitions. Resources, strategies, and behaviors are then compared across coalitions to gain insight into the degree of conflict and collaboration, and the distribution of power within a subsystem. This helps in classifying the subsystems (as collaborative, unitary, or conflictive, see below; see Nohrstedt and Weible 2010; Weible 2008), and adds to our understanding of policy change (Nohrstedt 2008; Sabatier and Weible 2007). Most often, already the mere number of coalitions per subsystem gives some important insight on policy process dynamics (Weible, Sabatier, and McQueen 2009; Pierce et al. 2017b).

Finally, the ACF unveils the characteristics and dynamics of *policy learning and policy change*. There are different pathways to policy learning and change: the first most often is defined as adaptation of belief systems to changing situations or external shocks. To explain policy change, the ACF emphasizes different pathways, including shifts in power balances (distribution of resources) between the dominant and minority coalitions, new actors integrating coalitions and the subsystem, altered interactions within and across coalitions, or brokers seeking policy stability. The underlying logic of learning and change is procedural. Studies of learning and change ideally investigate a process and thereby adopt a longitudinal perspective. In consequence, and at least implicitly, they compare subsystems and their characteristics over time (e.g. policy output at t_1 and t_2 in order to conceive policy change).

Different units of analysis follow these comparative elements (beliefs, resources, strategies, coordination, degree of conflict, coalitions, policy learning and policy change). If researchers compare beliefs, resources, or strategies among different cases, then the unit of analysis is typically the actor or the coalition. If one is interested in coordination or conflict among actors, the unit of analysis can either be the coalition or the subsystem. And finally comparative research about policy learning and policy change can be facilitated by comparing subsystems. But what

are the different cases that can be compared? Below we discuss how to design comparative policy process research in general, and through the ACF in particular.

3. COMPARATIVE ACF RESEARCH: GUIDELINES, EXPERIENCES AND SHORTCOMINGS

The vast majority of ACF studies have been single case studies; however, many of these case studies contain clear elements of comparison. Any effort to understand policy subsystems generally requires some contextual knowledge about actors, events, and policy content. This knowledge is not necessarily used as empirical evidence for hypothesis-testing, but is still very important for establishing an understanding of the level and intensity of policy conflict and what constitute the most important policy decisions, actors, and venues (Nohrstedt and Olofsson 2016b). This step usually involves some element of implicit comparison with other cases. Depending on the territorial scope of the subsystem under investigation (whether it is local, regional, national, or perhaps multilevel), the analyst can gain important insights by contrasting its features with other similar subsystems in other settings. Charles Ragin (1987) has classified such efforts to arrive at a more informed contextual understanding of single cases as comparative (see Dodds 2018).

The first steps in comparative research consist of knowing what to compare, what elements or variables to vary and what to keep constant to arrive at some general or even generalizable conclusions. In this way, comparative ACF applications liken to other comparative work in policy studies. There are three prominent ways to compare: over issues, over time, and across territorial units or jurisdictions. In ACF terms, the first (issue-related) comparative approach usually translates in the study of different subsystem topics (e.g. health and education, or energy and climate) where the subsystems belong to the same territorial unit or jurisdiction (e.g. within the same country, state, or catchment area). In longitudinal analysis, the same coalition, policy, or subsystem is analyzed at different points in time. And probably most challenging are cross-jurisdictional comparisons where the subsystem issue is kept constant, but where the territorial units change (different states, countries, etc.). This means that besides subsystem dynamics, also institutions of the political system can be drivers for coalition formation, policy learning or policy change, which might, from a research design point of view, consist in a challenge in clearly disentangling different (independent) variables.

3.1 Longitudinal and Cross-Sectional ACF Research

In comparison across time, there are several objectives for engaging in longitudinal comparison within the ACF, which are ultimately linked to testing of its hypotheses within each area of theoretical emphasis. In fact, all hypotheses that have been tested most frequently entail assumptions that necessitate comparative approaches to describe and explain changes in policy, coalitions, or beliefs through time (Weible, Sabatier, and McQueen 2009). These hypotheses (summarized in Table 5.1) concern the three areas of theoretical emphases within the ACF: policy change, stability of coalitions, and learning within or between coalitions. Testing each one of the hypotheses concerned with these areas requires some element of comparison of policy, coalitions, or belief systems through time.

Table 5.1 Empirical applications sorted by area of theoretical emphasis and subjects of longitudinal comparison

ACF areas of emphasis (times tested, percentage)	Subjects of longitudinal comparative analysis
Policy change (67, 42%)	Changes or stability in policy content – policy core and secondary aspects of policy programs
Coalition stability (32, 20%)	Stability or instability of coalition members (e.g. new affiliates or defectors), network relations, resources, and behavior
Learning (48, 29%)	Changes in actors’ belief systems, most likely policy core and secondary aspects

Note: This illustration is based on a meta-analysis by Pierce et al. (2017b) of 161 empirical applications of the ACF. Note that each area of emphasis contains several hypotheses (for a description of ACF’s hypotheses, see Jenkins-Smith et al. 2017).

3.2 Comparison Within and Across Countries

The ACF was originally developed to describe and explain policy processes in the United States and initially most empirical applications involved case studies of policy subsystems in environmental and energy policy issues. At the same time it was argued that the framework was “applicable to a variety of policy domains and political systems” (Sabatier and Jenkins-Smith 1999: 125). Yet critics questioned the generalizability of the framework beyond the pluralistic political system in the United States, maintaining that it would not work well as an organizing framework for understanding policy processes in other (less pluralistic) political systems. Questions were also raised concerning the portability of the framework to other policy domains outside environmental and energy policy issues, such as social policy. Responding to these critiques, Sabatier (1998) encouraged additional applications in other cases and contexts to explore the potential of the framework in different settings.

Although the ACF was created to describe and explain coalition behavior, policy change, and learning in individual policy subsystems, some of its central assumptions and extensions are suitable for comparative examination. Yet, the hypotheses of the ACF do not incorporate any assumptions about structural influences that we know vary across political systems and expect would influence policy subsystem affairs. For example, the framework does not state any hypotheses concerning the role of relatively stable parameters (attributes of the problem area, socio-cultural values, and constitutional structure) in shaping the structure and behavior of advocacy coalitions.

ACF scholars nevertheless have taken steps to encourage and promote examination of these factors across different political systems. Some of these theoretical specifications were a reaction to the critique regarding the applicability of the framework outside the United States. Addressing this issue, Daniel Kübler (2001) found evidence that coalition behavior is strongly influenced by political opportunity structures, including the level of territorial decentralization, separation of power, and institutionalization of procedures for direct democracy. Informed by this study, the notion of “long-term opportunity structure” was eventually added to the ACF to highlight the importance of the degree of consensus and openness of political systems. Specifically, Sabatier and Weible (2007) theorized that these system attributes would affect both the structure and behavior of advocacy coalitions. They suggested that “the higher the degree of consensus required, the more incentive coalitions have to be inclusive (rather than exclusive), to seek compromise and share information with opponents, and generally to

minimize devil shift”. Concerning openness, they proposed that advocacy coalitions are likely to have fewer actors in corporatist regimes, which are also likely to “create incentives for moderates to broker deals across coalitions” (Sabatier and Weible 2007: 200). Thus far though, no empirical study has yet explicitly tested these assumptions by comparing policy subsystems in different political systems.

In the past decades, there has been an explosion of applications of the ACF worldwide with hundreds of case studies from different countries, continents, and policy issues (Jenkins-Smith et al. 2014, 2017). A recent review of ACF studies globally in the period 2007–2014 found 161 applications covering a total of 54 countries (Pierce et al. 2017b).¹ If compared with an earlier review of studies published in the period of 1987–2006 identifying 80 applications (Weible, Sabatier, and McQueen 2009), the total number of applications written in English exceeds 240 and is probably closer to 500 if we include applications in different languages.

Among the 161 studies reported by Pierce et al. (2017a), 18 (11%) compared policy subsystems across countries. For the years 2014 until today, we conducted a supplementary search based on the same criteria as the ones adopted by Pierce et al. (2017a).² From the 94 identified ACF applications between fall/winter 2014 and spring/summer 2018,³ 23 adopted a comparative design either over time, subsystems or countries. Table 5.2 summarizes all the 28 cross-country comparisons between 2009 and today. Most studies compare countries within the same continent while 10 of them compared across continents, such as European countries with the US (e.g. Montpetit 2011) or European countries with countries in Africa and Asia (e.g. Huntjens et al. 2011).

The studies listed in Table 5.2 unveil some interesting trends regarding the comparative research agenda within the ACF. First, these 28 applications are consistent with the overall emphasis within the ACF on energy and environmental policy issues, yet the substantive breadth is expanding (see also Pierce et al. 2017b; Weible, Sabatier, and McQueen 2009). Second, almost all comparative studies focus on only one policy issue, which should help in reducing the number of potential rival explanations associated with subsystem-specific attributes. Third, these studies have been organized around different clusters of countries enabling comparisons of e.g. Western liberal democracies, former Eastern European countries, and combinations of these. Another and perhaps more unexpected cluster includes applications from Africa and Asia, and combinations of these and European countries.

There are also examples of case studies within the ACF that compare phenomena across policy subsystems within or across substantive policy issues within the same country. Some examples of these include the assessment of the capacities, activities, and interactions among proponents and opponents of hydraulic fracturing policy in New York, Colorado, and Texas (Weible and Heikkila 2016), learning within 10 collaborative partnerships in marine aquaculture subsystems on the east, west and gulf coasts of the United States (Leach et al. 2014), and the dynamics of coalition formation across three regional land use planning processes in California (Henry, Lubell, and McCoy 2011). What these studies show is that policy subsystems can vary greatly within the same political system.

These applications are clearly not sufficient as a basis for confirming or disconfirming the descriptive or explanatory validity of the ACF across different political systems. Nevertheless they showcase the breadth of comparative applications of the framework around the world. Although contributions from Europe and North America still dominate in comparative work, 14 out of the 40 countries in Table 5.2 (35%) are located in other parts of the world. In this

Table 5.2 List of cross-country comparative studies applying the ACF

Study (alphabetical order)	Policy issue(s)	Countries included
<i>Aamodt and Stensdal (2017)</i>	<i>Climate policy</i>	<i>Brazil, China, India</i>
Adams et al. (2014)	Spatial planning	Estonia, Latvia, Lithuania
Amougou and Larson (2008)	Internet diffusion	United States, France
Bandelow and Kundolf (2011)	Space policy	United Kingdom, Germany, the Netherlands, Sweden, France, Italy, Spain, Switzerland
Blatter (2009)	Environmental regulation	Germany, Switzerland, Austria
Brusis (2010)	Regional development	Bulgaria, Czech Republic, Serbia
<i>Cairney et al. (2018)</i>	<i>Fracking regulation</i>	<i>United Kingdom, Switzerland</i>
Cent et al. (2013)	Environmental regulation	Hungary, Poland
Cherlet and Venot (2013)	Water policy	Burkina Faso, Mali
<i>Daniell et al. (2014)</i>	<i>Water policy</i>	<i>Australia, China, Bulgaria</i>
Fidelman et al. (2014)	Marine policy	Indonesia, Malaysia, Philippines, Timor Leste, Papua New Guinea, Solomon Islands
<i>Gralepois et al. (2016)</i>	<i>Flood defense</i>	<i>Six European countries</i>
<i>Hughes and Meckling (2017)</i>	<i>Energy policy</i>	<i>United States, China</i>
Huntjens et al. (2011)	Water policy	Portugal, Ukraine, Hungary, the Netherlands, Czech Republic, Uzbekistan, Uganda, Tanzania, Rwanda, Burkina Faso, South Africa
<i>Ingold et al. (2017)</i>	<i>Fracking regulation</i>	<i>United Kingdom, Switzerland</i>
<i>Kukkonen et al. (2018)</i>	<i>Climate policy</i>	<i>Canada, United States, Brazil, India</i>
Mann and Gennaio (2010)	Environmental policy	United States, Switzerland
Meijerink (2008)	Marine policy	Belgium, the Netherlands
<i>Mockshell and Birner (2015)</i>	<i>Agri-food regulation</i>	<i>Ghana, Uganda</i>
Montpetit (2009)	Biotechnology policy	Canada, United States, France, United Kingdom, Belgium
Montpetit (2011)	Biotechnology policy	Canada, United States, France, United Kingdom
Montpetit (2012)	Biotechnology policy	Canada, United States, France, United Kingdom
Nedergaard (2009)	Multiple	Denmark, Finland, Iceland, Norway, Sweden
Sloboda et al. (2010)	Language policy	United Kingdom, Czech Republic, Hungary
<i>Sotirov and Winkel (2016)</i>	<i>Forest policy</i>	<i>Germany, Bulgaria</i>
Szarka (2010)	Wind power policy	Denmark, France, Spain, United Kingdom, Germany
<i>Valdes Coteria and Flores Crespo (2015)</i>	<i>Education policy</i>	<i>Brazil, Mexico</i>
Winkel and Sotirov (2011)	Forest policy	Germany, Bulgaria

Note: Studies identified by Pierce et al. (2017a) as well as own search for the years 2014–2018 applying the Pierce et al. (2017a) criteria. Search results by the authors in italics.

regard, comparative studies have gone further than the average of all ACF applications to explore countries outside Europe and North America (see Pierce et al. 2017b).

In addition to these comparative studies, there are a handful of other projects that in different ways have used the ACF as the organizing framework for comparing policy processes across countries. One recent illustrative example involves efforts to compare the same policy issue – unconventional oil and gas development (hydraulic fracturing, or so-called “fracking”) – across seven countries: the United States, Canada, Germany, France, Sweden, Switzerland, and Germany (Weible et al. 2019). Other studies have extended this comparative effort by developing and applying the same coding framework for analyzing news media material on hydraulic fracturing policy in Argentina, the United States, and China (Heikkila et al. 2019) as well as in France (Moysen et al. 2018) and the UK (Cairney, Fischer, and Ingold 2018). These

studies set out to answer the same set of research questions, targeting the attributes of advocacy coalitions and the status of public policy, within the same policy area across countries. Experiences from these case studies have then been collectively analyzed and synthesized by scholars from different countries (Cairney, Fischer, and Ingold 2018; Weible et al. 2018). This research design lies close to an ideal application of the method of structured and focused comparison detailed by George and Bennett (2004).

Other examples of less structured but still comparative efforts exploring the assumptions of the ACF involve special issues with case contributions from different countries. One of these, published in the *Policy Studies Journal* in 2011, contrasted different policy issues in carnivore management (Matti and Sandström 2011) and signals intelligence policy (Nohrstedt 2011) in Sweden, regional planning (Henry 2011) and foreign policy (Pierce 2011) subsystems in the United States, biotechnology policy in Canada (Montpetit 2011), and flood management policy in Hungary (Albright 2011). Another special issue, in the *Journal of Comparative Policy Analysis* in 2014, compared environmental policy in China (Han, Swedlow and Unger 2014), indigenous peoples' rights in the Philippines (Montefrio 2014), nuclear energy and forest management in India (Gupta 2014), and biosafety regulation in Kenya (Kingiri 2014).

One final group of comparative studies within the ACF include meta-reviews of previous empirical applications globally (Pierce et al. 2017a, 2017b; Weible, Sabatier, and McQueen 2009; Wellstead 2017), in specific countries, including China (Li and Weible 2018), South Korea (Park and Weible 2018), and Sweden (Nohrstedt and Olofsson 2016b), and within specific policy areas such as natural resource management (Sotirov and Memmler 2012). These reviews share an ambition to take stock of the available empirical evidence within the ACF – documenting for instance variability in application across policy issues, hypotheses tested, and application methods – and to identify areas to which ACF scholars have paid less attention.

Together these different case studies and comparative efforts provide a rich empirical base for making inferences about policy processes around the world. This is probably one of the richest and most well-documented sources of policy processes available within the field of public policy. Next, we turn to a discussion of some of the insights and lessons that have emerged from this material. Hereafter, we identify some remaining gaps and challenges for further advancing a comparative research agenda within the ACF research program.

4. INSIGHTS AND LESSONS

Our discussion of the insights and lessons that have emerged from the ACF research program is structured according to basic rationales in comparative politics and public policy. These can be summarized in four major reasons for engaging in comparison (Landman 2002). *Contextual description* aims at describing political phenomena and events in a particular country, group of countries, or within subunits with the goal to increase descriptive knowledge of the system under investigation. By engaging in *conceptual classification*, the comparativist establishes conceptual classifications to group countries, systems, or events into distinct categories with identifiable and shared characteristics. Next, *hypothesis-testing* involves efforts to search for factors that can explain what has been described or classified, including efforts to rule out rival explanations, as a basis for building theory. Finally, *prediction* is the logical extension of hypothesis-testing and involves attempts to foresee, in probabilistic terms, future political or policy outcomes based on generalizations from initial comparisons. Below we rely on these

rationales to structure our assessment of comparative analysis through the ACF. The reviews below do not examine single case studies of the ACF.

4.1 Contextual Description

One of the common usages of the ACF involves efforts to apply its core concepts to depict policy processes and issues around the world. In this regard, scholars have frequently used the ACF as an “organizing framework” (Sabatier 1998) to simplify and describe policy subsystems, including the participating actors and key developments in policy (stability or instances of minor and major change). In fact, the number of studies using the concepts of the ACF to contextually describe policy issues is considerably larger than the number of studies that engage in hypothesis-testing. For example, Pierce et al. (2017b) found a total of 512 articles in peer-reviewed journals that made references to any of the origin or revision publications of the ACF and included any of its key concepts as key words. Similarly, in a review of ACF applications in Sweden, Nohrstedt and Olofsson (2016b) noted that the number of studies using the ACF doubled when more cursory applications were considered.⁴

A recurrent goal of comparative applications of the ACF has been to provide detailed insight into policy subsystems in different countries. This effort has been guided by relatively specific descriptive research questions, such as what actors have been participating regularly in subsystem affairs, what are the attributes of advocacy coalitions, and what have been the most important changes in policy. One general insight that has emerged from efforts to answer these questions is that the ACF appears to provide sufficient descriptive validity across different types of political systems. That is, scholars regularly confirm that the notions of policy subsystem and advocacy coalitions are useful to simplify the complexity of the policy process by ordering information about actors and key events in policy making.

This work has generated some interesting observations about policy subsystems and the structure and behavior of advocacy coalitions in different political systems. Although scholars always encounter analytical challenges associated with delineating the boundaries of subsystems, the experience thus far is that the subsystem concept (with its defining territorial and substantive boundaries and associated participants) provides a clear and useful starting point for structuring collection of data about policy processes. In this regard the ACF works well as a framework to simplify the complexity of the policy process. Previous contributions similarly confirm that the existence of coalitions is sufficiently well documented in many different countries and is thus beyond dispute (Weible et al. 2019). It has been shown that the vast majority of ACF applications (143 out of 161, 89%) identify one coalition or more regardless of political context and most of these (121, 75%) find evidence of two or more coalitions (Pierce et al. 2017b).

But despite these advances, the fact remains that relatively few studies have examined more systematically context-driven variation in the composition, activities, stability, and resources of advocacy coalitions. Hence, no (or at least very limited) comparative empirical evidence exists to confirm or disconfirm the aforementioned assumption that coalitions will differ sys-

tematically due to the country-specific nature of the coalition opportunity structures (Weible et al. 2018). Yet, some sporadic observations have been made, for example:

- In *South Korea* representatives of the state bureaucracy, political parties, and large business organizations are key participants in policy subsystems, while non-governmental organizations and civic participation have been more limited (Park and Weible 2018).
- In *China* there are examples of advocacy coalitions established by the government but also cases of competing coalitions led by non-governmental actors (particularly think tanks and private industry) challenging government policy positions (Li and Weible 2018).
- In *Sweden* advocacy coalitions typically involve representatives of interest groups, government agencies, political parties, business, regional authorities, research institutes, unions, and the media, yet there is no evidence of participation of new organized participants such as policy professionals and think tanks (Nohrstedt and Olofsson 2016b).

These are a few selected example observations that illustrate the descriptive contextual properties of advocacy coalitions in different political systems, yet this is an area where scholars can clearly do more to advance the understanding of advocacy coalitions in different political contexts.

4.2 Conceptual Classification

In the ACF conceptual classification has evolved iteratively alongside contextual description. That is, efforts by scholars to describe policy subsystems in different countries and settings have generated new insights and suggestions regarding similarities and differences in the attributes of policy subsystems and advocacy coalitions in varying political systems. Below we discuss some detailed suggestions for conceptual classification that have emerged from these efforts. It should be noted that the topics below are not exhaustive – these rather represent some illustrative examples of conceptual classification in the ACF that have been advanced through comparative research.

Unitary, collaborative, and adversarial policy subsystems. While the early applications of the ACF focused on interactions among policy actors in high conflict situations, later applications have disaggregated different types of subsystems contexts according to different levels of conflict. This is partially a recognition that the level of conflict among policy actors is not static but may fluctuate over time. Based on this insight, Weible (2008) proposed three simplified subsystem types, distinguished by varying levels of conflict. Adversarial subsystems involve high levels of conflict among two or more competitive coalitions with low levels of inter-coalition belief compatibility. In collaborative subsystems, conflict remains at intermediate levels as the members of different coalitions share some common beliefs (usually confined to secondary aspects) with cross-coalition coordination. Lastly, unitary policy subsystems involve a single dominant advocacy coalition with substantial belief compatibility and coordination among its members. The classification of different subsystem types within the ACF has been empirically explored in studies of the role of scientists and information (Weible, Sabatier, and Pattison 2010) and instrumental in supporting theoretical refinements, for instance by specifying different pathways to policy change (Nohrstedt and Weible 2010).

Nascent and mature subsystems. Another way to classify policy subsystems within the ACF has been to separate subsystems that have existed for an extended period of time (“mature”) from those that are in the early stages of forming (“nascent”) (Sabatier and Jenkins-Smith 1999).

Mature subsystems involve a set of participants that regard themselves as a semi-autonomous community, who share an expertise in a particular domain, and who have sought to influence that domain over time, along with agencies, interest groups, and research institutions with subunits specializing in the topic. The behavior and dynamics of mature subsystems have been shown to be quite different from more nascent ones where participation is more fluid, the political positions of actors less clear, and where interactions are more collaborative (Beverwijk 2005; Fidelman et al. 2014; Stritch 2015). The classification of nascent and mature subsystems has been important for understanding coalition formation and behavior in relatively unstable political systems as well as in new policy issues, such as hydraulic fracturing (Ingold, Fischer, and Cairney 2017; Nohrstedt and Olofsson 2016a).

Principal and auxiliary coalition members. One of the major analytical challenges facing ACF analysts has been to categorize different types of coalition members. This partially stems from the theoretical definition of an advocacy coalition as consisting of policy actors that share beliefs and coordinate their behavior. In response, the classification into principal and auxiliary coalition members emerged from the insight that it is unrealistic to assume that all coalition members coordinate with each other (Zafonte and Sabatier 1998). Instead, one is likely to find that some members are more central, in the sense that they provide leadership for their coalition and facilitate coordination among most other members through time. These are actors that serve an entrepreneurial role for the coalition by connecting with most other members. In contrast, auxiliary members are more peripheral to the coalition and may only coordinate with some other members. These actors may lack sufficient resources to participate in coalition activities and, unlike principal members, may not view subsystem issues as salient to their beliefs. Distinguishing principal and auxiliary coalition members has been important to advance insights about the composition and behavior of coalitions, particularly their ability to achieve coordinated actions.

4.3 Hypothesis-Testing

Each one of the three areas of theoretical emphases within the ACF – coalitions, learning, and policy change – seeks to advance explanations of various phenomena in the policy process. These explanatory ambitions connect with some of the long-standing puzzles of public policy, focusing for instance on why advocacy coalitions form, why policy actors alter their beliefs through time, and under what conditions public policy changes. As mentioned above, the ACF offers a battery of testable hypotheses addressing the three topics of coalitions, learning, and policy change. In this regard, the long-term goal of the ACF is to help build theory about these phenomena. To date, these efforts have generated mixed results and not all hypotheses have been subject to empirical testing across multiple cases. For example, while hypotheses about the stability and change of policy and coalitions have been regularly examined by scholars, hypotheses about coordination have so far attracted less attention in empirical research (Weible, Sabatier, and McQueen 2009). We speculate that some of the ACF's hypotheses are more difficult to empirically test than others and also that scholars are generally more interested in some topics than others, which may explain these differences in application.

Given the focus of this chapter on comparative dimensions, we turn the attention to what role comparative methods have played within the ACF to advance these explanatory ambitions, focusing specifically on cross-country comparisons. To what extent and how have scholars relied upon comparative methods to test the assumptions of the ACF? Figure 5.1

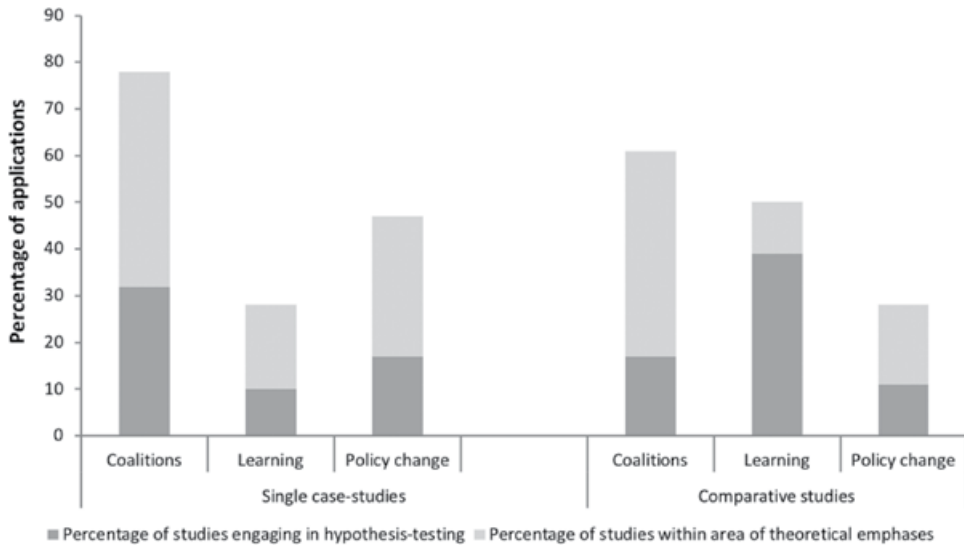


Figure 5.1 *Theoretical emphases and hypothesis-testing in empirical applications by type of study*

provides a summary of the share of applications comparing across countries versus single country studies across the three areas of theoretical emphases (data from Pierce et al. 2017b).

When considering the distribution of applications across areas of theoretical emphases, it can be noted, first, that ACF's assumptions about coalitions have received less attention in comparative work compared to single case-study applications. Specifically, whereas 78% of all case studies (112 out of 143) address questions about coalitions, this is a topic in more than half (61%, 11 out of 18) of the comparative studies. Second, the order is reversed when it comes to studying topics related to policy-oriented learning; learning is a topic covered by 50% of the comparative studies (9 out of 18) but only by 28% of the single case studies (40 out of 143). Finally, policy change is a less common topic in comparative studies (28%, 5 out of 18) than in single case studies (47%, 67 out of 143).

A slightly different picture emerges when looking more closely at the extent to which studies actually engage in hypothesis-testing. To do so, we focus on the percentage of studies within each category of studies (case studies and comparative studies, respectively) that engage with hypothesis-testing. When all areas of emphases are considered, hypothesis-testing is almost as common in single case-study applications (41%, 58 out of 143) as in comparative studies (44%, 8 out of 18). In this regard, both types of study designs have been important in driving theory development within the framework, yet the absolute number of single case studies ($n=58$) engaging in hypothesis-testing is obviously considerably larger than the number of hypothesis-testing comparative studies ($n=8$). Furthermore, the specific types of hypotheses tested have differed across different types of research designs. As shown in Figure 5.1, hypotheses about coalitions are more commonly tested in single case-studies (32%) compared to comparative studies (17%). In contrast, hypotheses related to the theoretical emphases on learning are a more common focus in comparative studies (39%) compared to single case

studies (10%). Finally, hypotheses about policy change are the least tested in the whole sample of applications, with 17% in single case studies and 11% in comparative studies.

It can also be noted that the number of areas of theoretical emphases covered by these tests are almost identical across single case studies and comparative studies (numbers not shown in Figure 5.1). Among the single case studies, only six (4%) test any combination of hypotheses covering all three areas of theoretical emphases and 16 (11%) test hypotheses from two areas. This is similar to the sample of comparative studies where 5% test hypotheses from all three areas and 11% test hypotheses from two areas.

Although these observations derive from a selection of ACF applications up to 2014, some general reflections can be made. Overall, the fact that we see relatively few applications testing hypotheses across several areas of theoretical emphases is an illustration of how the framework is being used to answer specific research questions about the policy process. This pattern supports the view of the ACF as a research program where scholars contribute to theory development by engaging in specific theoretical issues. Taking the step from single to multiple areas of emphases within the boundaries of one single study, regardless of research design, is very demanding since it increases the workload to compile data on different aspects of the policy process. At the same time, in some cases this might be desirable since testing of some of the hypotheses within the ACF requires attention to several areas of emphases. For example, the framework assumes that policy change depends on the behavior and interactions of advocacy coalitions, which means that data has to be collected for both coalitions and policy change.

We also note that given the analytical challenges associated with identifying coalitions, it is not surprising that coalitions is a more common theoretical emphasis in single case studies than in comparative studies. Assessing ACF's hypotheses about coalitions is generally demanding as it necessitates collection of information about actors, their beliefs, resources, and coordination behavior. In addition, one might be interested in studying changes in the composition or behavior of coalitions, which require collection of data at multiple points in time. Repeating this procedure for several cases can thus be overwhelming. However, there is no single way to measure coalitions within the ACF; analysts may employ different analytical strategies to identify coalitions, their actors, beliefs, and levels of coordination, which ultimately depend on the theoretical interests of the researcher (see Weible et al. 2019).

Learning is another concept in the ACF that has proven difficult to measure. Therefore, it is somewhat surprising that hypotheses about learning have been tested more frequently in comparative studies than in single case studies. The ACF defines policy-oriented learning as "relatively enduring alterations of thought or behavioral intentions that result from experience and/or new information and that are concerned with the attainment or revision of policy objectives" (Sabatier and Jenkins-Smith 1993: 123). Some comparative studies do not apply this definition systematically but rather make relatively superficial observations about changes in actors' beliefs, which are then taken as evidence of learning. Other studies (e.g. Montpetit 2011) examine learning in a more systematic fashion. Thus, given the central place of learning within the ACF and the attention it has received in comparative work, there seems to be a need for a concerted effort among ACF scholars to clarify conceptualization and measurement of various processes and products of learning in different political systems (Henry 2009, 2017; Jenkins-Smith et al. 2014; Moysen 2018).

Finally, we note that hypotheses about policy change have only been tested a few times in previous comparative work. This stands in contrast to the observation that ACF's hypotheses about external events and policy change are the most frequently tested hypotheses within the

research program. We speculate that several reasons account for this fact, some related to specificities in ACF research, others to the more enduring design challenges encountered in comparative research.

First, the lack of comparative examinations of policy change may be explained by the aforementioned difficulty to empirically validate the connection between coalition behavior and instances of policy change. Another potential explanation could be the lack of “best practices” for documenting policy change while accounting for context (Jenkins-Smith et al. 2014). However, recently scholars have engaged in efforts to remedy these limitations by attempts to document and explain variations in policy change and stability across several countries (Weible et al. 2019). Difficulties in how to account for different factors belonging to different “levels” (political system versus subsystem, see Weible et al. 2018) in comparative research may be another reason for the paucity of comparative studies about policy change.

When explaining policy change under the ACF, authors most often search for political drivers. In situations of minor policy change where secondary aspects are challenged, typical pathways for change are alterations in the power and coordination within and across coalitions, the presence of brokers, or policy-oriented learning (Sabatier and Weible 2007). Typically the ACF would direct researchers to analyze the politics over public policy issues at the scale of a policy subsystem. When comparing policy change of the same subsystem but in different countries or jurisdictions, a complex interplay between political system and subsystem dynamics starts to challenge conceptual and empirical research. A broad literature exists about how macro-institutional factors impact politics and policy (Gallagher, Laver, and Mair 1995; Lijphart 1984, 1999). Yet, relatively little systematic comparative research exists about the interplay between macro-institutions and politics on one side, and macro-institutions and policy change on the other (for exceptions see e.g. Cairney, Fischer, and Ingold 2018; Weible et al. 2018).

We acknowledge that degree of consensus versus pluralism in democracy has an impact on how conflictive or collaborative the set-up of coalitions within a subsystem is, and what the chances for brokerage are. And some political systems regularly produce more radical policies and policy changes than others. The ACF accounts for some institutional and contextual factors inside and outside the policy subsystem, including political system and subsystem opportunity structures that can have an impact on coalition formation and development. Internal and external shocks and focusing events can be major drivers for policy change; this finding appears mainly when comparative researchers study more radical instances of policy change. But when political system factors and when subsystem factors account for policy change, and how the two interact with each other, still has to be analyzed in greater detail.

4.4 Prediction

As Landman (2002) suggests, prediction is a logical extension of hypothesis-testing but the most difficult objective in comparative research. Prediction (in terms of attempts to foresee outcomes in a probabilistic sense in other countries or systems based on generalizations from comparative research) has not been an explicit objective within the ACF research program. There is not a single study or application of the ACF that claims to use comparative work to make predictions about present or future policies or outcomes in other countries. That said, we can make a few observations about the ACF in relation to predictive ambitions in public policy.

One thing that distinguishes the ACF from some other comparative public policy research is the ambition to blend descriptive and explanatory ambitions. Generally the goal of comparative public policy research varies from understanding countries' selection and usage of different mixes of policy instruments (e.g. Dodds 2018) to understanding how cultural, economic, political, and institutional context enables and constrains the adoption of certain policies (e.g. Adolino and Blake 2001). The ACF adds to these ambitions by specifying a range of assumptions concerning the structure and behavior of advocacy coalitions in specialized policy subsystems. Some of these assumptions are framed in probabilistic terms, for example that on major controversies, advocacy coalitions are likely to remain stable over time. Thus, with respect to political arrangements on policy topics, the expectation of the ACF that advocacy coalitions exist under certain conditions has been confirmed. The probabilistic prediction that certain pathways to policy change matter and possibly vary in some contexts rather than others has not been confirmed or disconfirmed. Jang, Weible, and Park (2016) cautiously observed that the South Korean political system might be prone to some pathways to change (e.g. changes in governing coalitions and events from outside the country) rather than in other political systems as might be found in the United States. Ingold and Varone (2012) found evidence that the Swiss consensus- and direct-democratic system is particularly open for collaboration across coalitions and the presence of policy brokers seeking stability and compromise. With the aim to test the so-called "belief homophily hypothesis", the hypothesis that two actors sharing similar beliefs tend to coordinate action, work by Calanni et al. (2015) and Ingold and Fischer (2014) leads us to conclude that the degree of conflict within the subsystem matters; in conflictive subsystems, belief homophily seems to be a strong predictor of actor coordination, whereas in collaborative subsystems, actors have a tendency to collaborate with powerful organizations rather than (and sometimes in addition to) ideological peers (Henry 2011). All these rather explorative empirical evidences could lead to some future predictions. And if we consider prediction probabilistically then there is a strong argument that the ACF could be useful in descriptive and explanatory understanding by different contexts if we maintain and build upon its comparative traditions.

5. CONCLUSION

The ACF remains one of the most widely applied frameworks of the policy process. In the recent decades empirical applications of the framework have expanded to new countries and settings. Experiences emerging from these applications have contributed to healthy debate among scholars concerning its descriptive and explanatory validity under different cultural, political, and economic conditions. This debate has been propelled by scholars who ask fundamental and critical questions about the basic assumptions of the framework, focusing on the nature and evolution of policy subsystems, the structure and behavior of advocacy coalitions, and the drivers of policy-oriented learning and policy change.

These contributions illustrate the merits of the ACF as a research program, which provides common vocabulary to support ongoing dialogue and exchange of experience among scholars engaging with the same research questions. The ACF has hereby served as a platform for theory development, advancing new insights about the policy process. Yet, much of this research has evolved through implicit comparisons, essentially driven by interactions among scholars working with different projects involving different cases, empirical data, and analyti-

cal methodologies. Although we have noted growing interest and engagement in comparative research across countries, there is still considerable unexploited potential within the ACF program to do more. In closing, we identify a few issues that would require attention by scholars to collectively push a comparative research agenda within the ACF:

- *Think classes of cases.* Oftentimes our case-selection strategies are determined at the early stages of the research process, quite often already at the stage of writing grant applications. Also, the usual criterion for case selection is novelty – the ambition to unpack new and exciting phenomena that have not been studied in-depth before. We imagine that in some instances more can be done to pay closer attention to the potential for comparison with similar cases in other countries in order to support more systematic comparison and orderly accumulation of knowledge within the ACF. Hereby researchers can collectively engage in comparative efforts around the same “class” of cases, including attempts to answer the same questions or addressing the same aspects in relation to one phenomenon in different settings (George and Bennett 2004). Defining the universe of potential phenomena a priori to guide these investigations is difficult but may involve cases of the same substantive policy area (e.g. energy, environmental, or social policy) or the same aspect of the policy process (e.g. formation of advocacy coalition or instances of policy change).
- *Sustain conceptual consistency.* One major challenge in comparative research in general is to apply the same concepts consistently across units and avoid conceptual stretching, that is, the distortion of concepts to fit new cases (Sartori 1970). Applied to the ACF, this means that we would strive for similar usages of its core concepts – e.g. advocacy coalitions, learning, and policy change – when applied in different contexts. These aspirations, however, would have to be balanced against the need to discover nuances in policy processes across political systems as well as the demand for alternative data collection strategies in different settings. The ACF provides clear enough definitions to enable researchers to engage in comparative work across countries. In doing so, scholars should be attentive to contextual nuances. For example, members of advocacy coalitions are likely to utilize different strategies of coordination depending on specific institutional conditions. Thus, moving forward researchers should be careful to ensure that the concepts of the ACF are consistently defined and validly applied while allowing for contextual variations.
- *Tap into best practice methods.* When engaging in comparative work across countries scholars will encounter a number of barriers that make it difficult to repeat identical strategies for data collection and methods of analysis. For example, a survey might be feasible in one country or subsystem but not another. Any approach that is selected must be weighed against its potential costs (e.g. time, resources, potential biases, and risk for missing data). Thus, collecting data for ACF analysis involves making wise choices for ensuring reliability and validity. The ACF research program offers considerable experience to draw on concerning state of the art methods for studying policy subsystems across different settings and countries. Examples include using different sources of data for documenting beliefs (for instance legislative documents, news media, surveys, and interviews) and different versions of network analysis (for instance documenting coordination networks, information networks, and ally and opponent networks) (see Henry, Lubell, and McCoy 2012; Weible et al. 2019). Similarly, different methods for studying learning and policy changes have evolved over time and offer useful guidance to scholars interested in studying these topics.

- *Identify decisive systemic attributes.* One of the most intriguing questions concerning the applicability of the ACF across countries is what contextual attributes are important in shaping subsystem affairs. Concretely, what are the decisive characteristics of a polity that shape the composition and behavior of advocacy coalitions and processes of learning and policy change? This question has received growing attention following applications of the ACF in new countries and contexts. However, thus far studies have not done enough to empirically examine whether there are certain systemic attributes that seem to influence subsystems across countries. Coalition opportunity structures – including the openness of political systems and the consensus needed for major policy change – constituted an important addition to the ACF, yet there might be other attributes that are important as well, for instance partisan elite responsiveness, top-down political suppression of political behavior, and susceptibility to international influences (as external events) within any given country (Jang, Weible, and Park 2016; Li and Weible 2018; Nohrstedt 2010). Advancing knowledge on this front is clearly contingent upon the collective effort by scholars to explore these possibilities and others as well.
- *Take inspiration from other fields.* Although major steps have been taken within the ACF to advance a comparative research agenda, more work obviously remains. Our objective in this chapter has been to support the effort by identifying and discussing some of the key issues in comparative ACF research and potential strategies for addressing these. The community of ACF scholars provides experience and “best practices” for researchers to draw on, yet inspiration may also come from other related fields in comparative politics and comparative public administration. Tapping into the long-term experience (good and bad) from these fields is one way to support a comparative research agenda within the ACF into the future.

The ACF is a potentially powerful approach for conducting comparative public policy research. This chapter summarizes some of its strengths, particularly in establishing a research program for comparing descriptive and explanatory insights in the area of coalitions, learning, and policy change. This chapter also identifies some challenges in moving this research forward, particularly given the variability in how some concepts are operationalized and measured, inconsistency in testing hypotheses, and linking contextual factors to conceptual observations. Given these strengths and challenges, the ACF remains a promising framework for the field of public policy research.

NOTES

1. Criteria for inclusion as an ACF “application” included (1) citation of at least one of the ACF origin and revision publications; (2) mention in title and/or abstract of two or more key works associated with the ACF (e.g. learning, coalitions, advocacy); (3) analysis of empirical data; (4) focus on defined policy topic; (5) analysis of one or more theoretical emphases in the ACF (learning, policy change, coalitions) (Pierce et al. 2017b).
2. From the year 2014 we included the months October, November, and December and the ACF applications therein not yet included by Pierce et al. (2017a).
3. The first application included being Cho published in October 2014 and the last being Haukkala published in June 2018.

4. These “cursory” applications involve empirical studies that use some of the core concepts of the ACF – commonly subsystems, advocacy coalitions, policy change, and learning – for descriptive or explanatory purposes, without operationalizing these concepts in a systematic fashion.

REFERENCES

- Aamodt, S. and Stensdal, I. (2017). Seizing policy windows: policy influence of climate advocacy coalitions in Brazil, China, and India, 2000–2015. *Global Environmental Change* 46: 114–25.
- Adams, N., Cotella, G., and Nunes, R. (2014). The engagement of territorial knowledge communities with European spatial planning and the territorial cohesion debate: a Baltic perspective. *European Planning Studies* 22 (4): 712–34.
- Adolino, J. and Blake, C. (2001). *Comparing Public Policies: Issues and Choices in Six Industrialized Countries*. Washington, DC: CQ Press.
- Albright, E. (2011). Policy change and learning in response to extreme flood events in Hungary: an advocacy coalition approach. *Policy Studies Journal* 39 (3): 485–511.
- Amougou, J. and Larson, J. (2008). Comparing implementation of internet diffusion in the United States and France: policies, beliefs, and institutions. *Review of Policy Research* 25 (6): 563–78.
- Bandelow, N. and Kundolf, S. (2011). Belief systems and the emergence of advocacy coalitions in nascent subsystems: a case study of the European GNSS Program Galileo. *German Policy Studies* 7 (2): 113–39.
- Beverwijk, J. (2005). The genesis of a system: coalition formation in Mozambican higher education, 1993–2003. PhD dissertation, University of Twente.
- Blatter, J. (2009). Performing symbolic politics and international environmental regulation: tracing and theorizing a causal mechanism beyond regime theory. *Global Environmental Politics* 9 (4): 81–110.
- Brusis, M. (2010). European Union incentives and regional interest representation in Central and East European countries. *Acta Politica* 45 (1–2): 70–89.
- Cairney, P., Fischer, M., and Ingold, K. (2018). Fracking in the UK and Switzerland: why differences in policymaking systems don’t always produce different outputs and outcomes. *Policy & Politics* 46 (1): 125–47.
- Calanni, J., Siddiki, S., Weible, C., and Leach, W. (2015). Explaining coordination in collaborative partnerships and clarifying the scope of the belief homophily hypothesis. *Journal of Public Administration Research and Theory* 25 (3): 901–27.
- Cent, J., Mertens, C., and Niedzialkowski, K. (2013). Roles and impacts of non-governmental organizations in Natura 2000 implementation in Hungary and Poland. *Environmental Conservation* 40 (2): 119–28.
- Cherlet, J. and Venot, J.-P. (2013). Structure and agency: understanding water policy changes in West Africa. *Water Policy* 15 (3): 479–95.
- Cho, H. (2014). Advocacy coalition for historic preservation in the U.S.: changes in motivations. *Journal of Arts Management, Law, and Society* 44 (4): 234–45.
- Daniell, K., Coombes, P., and White, I. (2014). Politics of innovation in multi-level water governance systems. *Journal of Hydrology* 519: 2415–35.
- Dodds, A. (2018). *Comparative Public Policy* (2nd edition). London: Palgrave Macmillan.
- Fidelman, P., Evans, L., Foale, S., Weible, C., von Heland, F., and Elgin, D. (2014). Coalition cohesion for regional marine governance: a stakeholder analysis of the Coral Triangle initiative. *Ocean & Coastal Management* 95: 117–28.
- Gallagher, M., Laver, M., and Mair, P. (1995). *Representative Government in Modern Europe*. New York: McGraw-Hill.
- George, A. and Bennett, A. (2004). *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- Gralepois, M., Larrue, C., Wiering, M., Crabbé, A., Tapsell, S., Mees, H., Ek, K., and Szwed, M. (2016). Is flood defense changing in nature? Shifts in the flood defense strategy in six European countries. *Ecology and Society* 21 (4): 1–11.

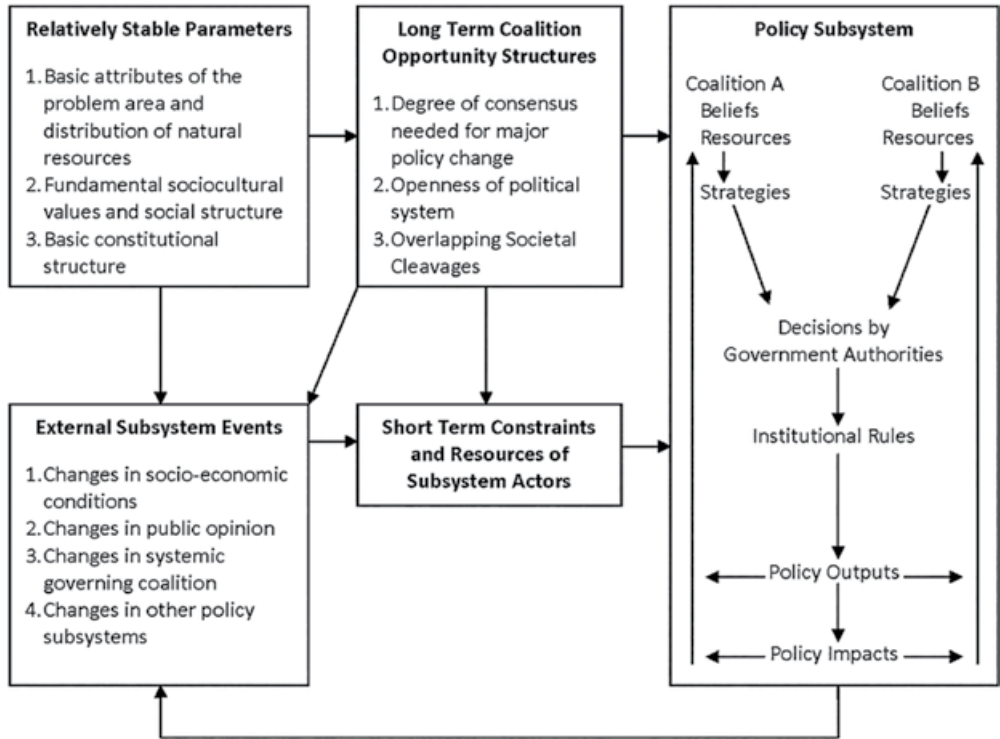
- Gupta, K. (2014). A comparative policy analysis of coalition strategies: case studies of nuclear energy and forest management in India. *Journal of Comparative Policy Analysis: Research and Practice* 16 (4): 356–72.
- Han, H., Swedlow, B., and Unger, D. (2014). Policy advocacy coalitions as causes of policy change in China? Analyzing evidence from contemporary environmental politics. *Journal of Comparative Policy Analysis: Research and Practice* 16 (4): 313–34.
- Haukkala, T. (2018). A struggle for change: the formation of a green-transition advocacy coalition in Finland. *Environmental Innovation and Societal Transitions* 27: 146–56.
- Heikkilä, T., Berardo, R., Weible, C. M., and Yi, H. (2019). A comparative view of advocacy coalitions: exploring shale development politics in the United States, Argentina, and China. *Journal of Comparative Policy Analysis: Research and Practice* 21 (2): 151–66.
- Henry, A. D. (2009). The challenge of learning for sustainability: a prolegomenon to theory. *Human Ecology Review* 16 (2): 131–40.
- Henry, A. D. (2011). Ideology, power, and the structure of policy networks. *Policy Studies Journal* 39 (3): 361–83.
- Henry, A. D. (2017). Network segregation and policy learning. In J. Victor, A. Montgomery, and M. Lubell (Eds.), *The Oxford Handbook of Political Networks*. Oxford: Oxford University Press, 559–87.
- Henry, A. D., Lubell, M., and McCoy, M. (2011). Belief systems and social capital as drivers of policy network structure: the case of California regional planning. *Journal of Public Administration Research and Theory* 21 (3): 419–44.
- Henry, A. D., Lubell, M., and McCoy, M. (2012). Survey-based measurement of public management and policy networks. *Journal of Policy Analysis and Management* 31 (2): 432–52.
- Hughes, L. and Meckling, J. (2017). The politics of renewable energy trade: the US–China solar dispute. *Energy Policy* 105: 256–62.
- Huntjens, P., Pahl-Wostl, C., Rihoux, B., Schlüter, M., Flachner, Z., Neto, S., Koskova, R., Dickens, C., and Kiti, I. N. (2011). Adaptive water management and policy learning in a changing climate: a formal comparative analysis of eight water management regimes in Europe, Africa, and Asia. *Environmental Policy and Governance* 21 (3): 145–63.
- Ingold, K. (2011). Network structures within policy processes: coalitions, power, and brokerage in Swiss climate policy. *Policy Studies Journal* 39 (3): 435–59.
- Ingold, K. and Fischer, M. (2014). Drivers of collaboration: what impact do joint preferences and actors' power have? An illustration of the Swiss climate policy over 15 years. *Global Environmental Change* 24: 88–98.
- Ingold, K., Fischer, M., and Cairney, P. (2017). Drivers for policy agreement in nascent subsystems: an application of the advocacy coalition framework to fracking policy in Switzerland and the UK. *Policy Studies Journal* 45 (3): 442–63.
- Ingold, K. and Varone, F. (2012). Treating policy brokers seriously: evidence from the climate policy. *Journal of Public Administration Research and Theory* 22 (2): 319–46.
- Jang, S., Weible, C. M., and Park, K. (2016). Policy processes in South Korea through the lens of the Advocacy Coalition Framework. *Journal of Asian Public Policy* 9 (3): 274–90.
- Jenkins-Smith, H. C., Nohrstedt, D., Weible, C. M., and Sabatier, P. A. (2014). The Advocacy Coalition Framework: foundations, evolution, and ongoing research. In P. A. Sabatier and C. M. Weible (Eds.), *Theories of the Policy Process* (3rd edition). Boulder, CO: Westview Press, 183–223.
- Jenkins-Smith, H. C., Nohrstedt, D., Weible, C. M., and Sabatier, P. A. (2017). The Advocacy Coalition Framework: an overview of the research program. In C. M. Weible and P. A. Sabatier (Eds.), *Theories of the Policy Process* (4th edition). Boulder, CO: Westview Press, 135–72.
- Kingiri, A. (2014). Comparative strategic behavior of advocacy coalitions and policy brokers: the case of Kenya's biosafety regulatory policy. *Journal of Comparative Policy Analysis: Research and Practice* 16 (4): 373–95.
- Kübler, D. (2001). Understanding policy change with the Advocacy Coalition Framework: an application to Swiss drug policy. *Journal of European Public Policy* 8 (4): 623–41.
- Kukkonen, A., Ylä-Anttila, T., Swarnakar, P., Broadbent, J., Lahsen, M., and Stoddart, M. (2018). International organizations, advocacy coalitions, and domestication of global norms: debates on climate change in Canada, the US, Brazil, and India. *Environmental Science & Policy* 81: 54–62.
- Landman, T. (2002). *Issues and Methods in Comparative Politics: An Introduction*. London: Routledge.

- Laudan, L. (1978). *Progress and Its Problems: Towards a Theory of Scientific Growth*. Berkeley, CA: University of California Press.
- Leach, W. D., Weible, C. M., Vince, S. R., Siddiki, S. N., and Calanni, J. (2014). Fostering learning through collaboration: knowledge acquisition and belief change in marine aquaculture partnerships. *Journal of Public Administration Research and Theory* 24 (3): 591–622.
- Li, W. and Weible, C. M. (2018). Understanding China's policy processes: a review of the Advocacy Coalition Framework's applications in China. Paper prepared for ECPR ACF workshop, Nicosia, Cyprus, April 11–13.
- Lijphart, A. (1984). *Democracies*. New Haven, CT: Yale University Press.
- Lijphart, A. (1999). *Patterns of Democracy*. New Haven, CT: Yale University Press.
- Mann, S. and Gennaio, M.-P. (2010). The central role of centralisation in environmental policy initialisation. *Journal of Environmental Planning and Management* 53 (3): 283–95.
- Matti, S. and Sandström, A. (2011). The rationale determining advocacy coalitions: examining coordination networks and corresponding beliefs. *Policy Studies Journal* 39 (3): 385–410.
- Meijerink, S. (2008). Explaining continuity and change in international policies: issue linkage, venue change, and learning on policies for the River Scheldt Estuary 1967–2005. *Environment and Planning A* 40 (4): 848–66.
- Mockshell, J. and Birner, R. (2015). Donors and domestic policy makers: two worlds in agricultural policy-making? *Food Policy* 55 (August): 1–14.
- Montefrío, M. (2014). State versus indigenous peoples' rights: comparative analysis of stable system parameters, policy constraints and the process of delegitimization. *Journal of Comparative Policy Analysis: Research and Practice* 16 (4): 335–55.
- Montpetit, É. (2009). Governance and policy learning in the European Union: a comparison with North America. *Journal of European Public Policy* 16 (8): 1185–203.
- Montpetit, É. (2011). Scientific credibility, disagreement, and error costs in 17 biotechnology policy subsystems. *Policy Studies Journal* 39 (3): 513–33.
- Montpetit, É. (2012). Does holding beliefs with conviction prevent policy actors from adopting a compromising attitude? *Political Studies* 60 (3): 621–42.
- Moyson, S. (2018). Policy learning over a decade or more and the role of interests therein: the European liberalization policy process of Belgian network industries. *Public Policy and Administration* 33 (1): 88–117.
- Moyson, S., Aubin, D., Chailleux, S., Fievet, B., and Plancq, M. (2018). Which first? Relations between policy learning and policy change in the French politics of hydraulic fracturing. Paper prepared for ECPR ACF workshop, Nicosia, Cyprus, April 11–13.
- Nedergaard, P. (2009). Policy learning processes in international committees: the case of the civil servant committees of the Nordic Council of Ministers. *Public Management Review* 11 (1): 23–37.
- Nohrstedt, D. (2008). The politics of crisis policymaking: Chernobyl and Swedish nuclear energy policy. *Policy Studies Journal* 36 (2): 257–78.
- Nohrstedt, D. (2010). Do advocacy coalitions matter? Crisis and change in Swedish nuclear energy policy. *Journal of Public Administration Research and Theory* 20 (2): 309–33.
- Nohrstedt, D. (2011). Shifting resources and venues producing policy change in contested subsystems: a case study of Swedish signals intelligence policy. *Policy Studies Journal* 39 (3): 461–84.
- Nohrstedt, D. and Olofsson, K. (2016a). Advocacy coalition politics and strategies on hydraulic fracturing in Sweden. In C. M. Weible, T. Heikkila, K. Ingold, and M. Fischer (Eds.), *Policy Debates on Hydraulic Fracturing*. New York: Palgrave Macmillan, 147–75.
- Nohrstedt, D. and Olofsson, K. (2016b). A review of applications of the Advocacy Coalition Framework in Swedish policy processes. *European Policy Analysis* 2 (2): 18–42.
- Nohrstedt, D. and Weible, C. M. (2010). The logic of policy change after crisis: proximity and subsystem interaction. *Risks, Hazards, and Crisis in Public Policy* 1 (2): 1–32.
- Ostrom, E. (2005). *Understanding Institutional Diversity*. Princeton, NJ: Princeton University Press.
- Park, K. and Weible, C. M. (2018). Developing policy theories in South Korea: lessons from the Advocacy Coalition Framework. *Journal of Asian Public Policy* 11 (2): 136–50.
- Pierce, J. J. (2011). Coalition stability and belief change: advocacy coalitions in U.S. foreign policy and the creation of Israel, 1922–44. *Policy Studies Journal* 39 (3): 411–34.

- Pierce, J. J., Hicks, K., Peterson, H. L., and Giordano, L. (2017a). Common approaches for studying the Advocacy Coalition Framework: review of methods and exemplary practices. Paper prepared for ECPR General Conference, Oslo, Norway, September 6–9.
- Pierce, J. J., Peterson, H. L., Jones, M. D., Garrard, S., and Vu, T. (2017b). There and back again: a tale of the Advocacy Coalition Framework. *Policy Studies Journal* 45 (S1): 13–46.
- Ragin, C. (1987). *The Comparative Method: Moving Beyond Qualitative and Quantitative Methods*. Berkeley, CA: University of California Press.
- Sabatier, P. A. (1986). Top-down and bottom-up models of policy implementation: a critical analysis and suggested synthesis. *Journal of Public Policy* 6 (January): 21–48.
- Sabatier, P. A. (1988). An advocacy coalition model of policy change and the role of policy-oriented learning therein. *Policy Sciences* 21: 129–68.
- Sabatier, P. A. (1998). The Advocacy Coalition Framework: revisions and relevance for Europe. *Journal of European Public Policy* 5: 98–130.
- Sabatier, P. A. and Jenkins-Smith, H. C. (1993). *Policy Change and Learning: An Advocacy Coalition Approach*. Boulder, CO: Westview Press.
- Sabatier, P. A. and Jenkins-Smith, H. C. (1999). The Advocacy Coalition Framework: an assessment. In P. A. Sabatier and H. C. Jenkins-Smith (Eds.), *Theories of the Policy Process*. Boulder, CO: Westview Press, 117–68.
- Sabatier, P. A. and Weible, C. M. (2007). The Advocacy Coalition Framework: innovations and clarifications. In P. Sabatier and C. Weible (Eds.), *Theories of the Policy Process* (2nd edition). Boulder, CO: Westview Press, 189–200.
- Sartori, G. (1970). Concept misformation in comparative politics. *American Political Science Review* 64 (4): 1033–53.
- Sloboda, M., Szabó-Gilinger, E., Vigers, D., and Šimičić, L. (2010). Carrying out a language policy change: advocacy coalitions and the management of the linguistic landscape. *Current Issues in Language Planning* 11 (2): 95–113.
- Sotirov, M. and Memmler, M. (2012). The Advocacy Coalition Framework in natural resource policy studies: recent experiences and further prospects. *Forest Policy and Economics* 16 (March): 51–64.
- Sotirov, M. and Winkel, G. (2016). Toward a cognitive theory of shifting coalitions and policy change: linking the Advocacy Coalition Framework and cultural theory. *Policy Sciences* 49 (2): 125–54.
- Stritch, A. (2015). The Advocacy Coalition Framework and nascent subsystems: trade union disclosure policy in Canada. *Policy Studies Journal* 43 (4): 437–55.
- Szarka, J. (2010). Bringing interests back in: using coalition theories to explain European wind power policies. *Journal of European Public Policy* 17 (6): 836–53.
- Valdes Cotera, R. and Flores Crespo, P. (2015). Analysis of public policies in basic education of young people and adults: a comparative analysis between Brazil and Mexico. *Revista Iberoamericana de Educacion* 69 (1): 61–78.
- Weible, C. M. (2008). Expert-based information and policy subsystems: a review and synthesis. *Policy Studies Journal* 36 (4): 615–35.
- Weible, C. M. and Heikkila, T. (2016). Comparing the politics of hydraulic fracturing in New York, Colorado, and Texas. *Review of Policy Research* 33 (3): 232–50.
- Weible, C. M., Ingold, K., Nohrstedt, D., Henry, A. D., and Jenkins-Smith, H. C. (2019). Sharpening advocacy coalitions. *Policy Studies Journal*. Early view (published online 28 June 2019).
- Weible, C. M., Nohrstedt, D., Ingold, K., Heikkila, T., Fischer, M., and Cairney, P. (2018). Advocacy coalitions and policy change: comparative insights from fracking politics across ten countries. Paper prepared for ECPR General Conference, Hamburg Universität, Germany, August 22–25.
- Weible, C. M. and Sabatier, P. A. (2005). Comparing policy networks: marine protected areas in California. *Policy Studies Journal* 33 (2): 181–201.
- Weible, C. M., Sabatier, P. A., Jenkins-Smith, H. C., Nohrstedt, D., and Henry, A. D. (2011). A quarter century of the Advocacy Coalition Framework: an introduction to the Special Issue. *Policy Studies Journal* 39 (3): 349–60.
- Weible, C. M., Sabatier, P. A., and McQueen, K. (2009). Themes and variations: taking stock of the Advocacy Coalition Framework. *Policy Studies Journal* 37 (1): 121–40.

- Weible, C. M., Sabatier, P. A., and Pattison, A. (2010). Harnessing expert-based information for learning and the sustainable management of complex socio-ecological systems. *Environmental Science and Policy* 13: 522–34.
- Wellstead, A. (2017). Plus ça change, plus c'est la même chose? A review of Paul Sabatier's "An advocacy coalition framework of policy change and the role of policy-oriented learning therein." *Policy Sciences* 50 (4): 549–61.
- Winkel, G. and Sotirov, M. (2011). An obituary for national forest programmes? Analyzing and learning from the strategic use of "new modes of governance" in Germany and Bulgaria. *Forest Policy and Economics* 13 (2): 143–54.
- Zafonte, M. and Sabatier, P. A. (1998). Shared beliefs and imposed interdependencies as determinants of ally networks in overlapping subsystems. *Journal of Theoretical Politics* 10 (4): 473–505.

APPENDIX



Source: Adapted from Weible et al. (2011).

Figure 5A.1 Flow diagram of the Advocacy Coalition Framework

6. Comparing agenda-settings: the Comparative Agendas Project

Laura Chaqués Bonafont, Christoffer Green-Pedersen and Henrik Bech Seeberg

1. INTRODUCTION

Studies of agenda setting have a long tradition within public policy research. Seminal work by Bachrach and Baratz (1962), Cobb and Elder (1983), and Kingdon (1984) has continued to prove its importance over decades. In recent years, this research tradition has taken a new turn and gained considerable momentum, not least because of the establishment of the Comparative Agendas Project (CAP) and the research based on the data collected in this project. The CAP is built on the foundational work by Baumgartner and Jones (1993) and their dataset on the US, and today, it brings together research groups of 18 different countries in Europe, Asia, and Latin America. The growth of the CAP dataset has generated a flourishing literature that has moved the policy agenda-setting tradition forward in several ways, both theoretically and in terms of methods. This has generated a much better understanding of agenda-setting dynamics across time, issues, and countries. Thus, agenda-setting processes around policy issues are in many ways better understood than just ten years ago.

This chapter provides an overview of the development of the policy agenda literature within the last decade focusing on the development of the CAP datasets (available through <https://www.comperativeagendas.net>). The chapter is structured as follows. The first section discusses previous contributions to the analysis of agenda setting, from the classical insights of policy agenda-setting literature to more recent theoretical developments, and the second section focuses on the punctuated equilibrium theory. The third section offers a general discussion about the contribution of the Comparative Agendas Project datasets to the study of agenda setting and describes several indicators used to analyze agenda dynamics (entropy, kurtosis, etc.). The last part of the chapter suggests next steps for future policy agenda-setting research.

2. AGENDA-SETTING MODELS

In its origins, agenda-setting studies focused mainly on the study of issue conflict and power. The so-called “conflict management approach”, which was the dominant approach during the 1960s and 1970s, emphasizes the problems associated with the functioning of the democratic system. The main claim of some of the major contributions during this period, such as Schattschneider (1960) or Bachrach and Baratz (1962), was that policy makers were neither paying attention to the issues that directly affected a larger part of the population, such as poverty and social inequality, nor responding to the issues that citizens identify as the most important problems for the nation.

These studies presented a critical voice to the pluralist view of politics. The problem of the pluralist approach is that most of the time, “the heavenly chorus sings with a strong upper class accent” (Schattschneider 1960: 35). In particular, Bachrach and Baratz (1962) emphasized that power may be, and often is, exercised by confining the scope of decision making to relatively “safe” issues. Most citizens cannot overcome the powerful barriers of entry that characterize the pressure system. Any political conflict is characterized by the existence of privileged groups, whose main goal is to maintain the status quo, keeping “outsiders”, and non-privileged groups, whose goal it is to expand political conflict and enter into the decision-making process out of the political game.

After Bachrach and Baratz (1962) and Schattschneider (1960), authors such as Cobb and Elder (1983) focus the analysis on policy actors’ ability to transform the values and beliefs about the nature of policy problems and/or policy solutions. Persuasion and argumentation together with the use of material resources increase the chances of policy actors – including both advocates of policy change and defendants of the status quo – to influence policy decisions. Following Schattschneider’s (1960) observation about issue conflict expansion, Cobb and Elder (1971) highlight that issues arise out of group conflict and reach decision-makers when the issue is expanded through persuasion and argumentation in terms of its scope, potency, and proximity. Issues that are broader in scope – that is, more ambiguous, less technical, of higher social significance, and of longer temporal relevance – are more likely to reach the agenda than others.

In contrast to the conflict management perspective, John Kingdon (1984) developed the multiple stream model of agenda setting by adopting elements from organizational theory and evolutionary biology. As John et al. (2013) emphasize, Kingdon’s model rejects the idea of agenda setting as a linear process in which a policy maker identifies a problem to solve, the bureaucracy produces a range of possible solutions, and the policy maker selects the best choice. Rather, Kingdon uses the metaphor of three separate “streams” to describe the gap between policy makers’ attention to a problem and their adoption of a meaningful solution.

Kingdon’s policy window model suggests that policy changes are a result of a complex interaction of unrelated streams – problem, policy, and political streams – that come together quite unpredictably. A policy window opens either by the appearance of compelling problems or by happenings in the political stream. Policy entrepreneurs play a key role to couple the separate streams together: “They keep their proposal ready, waiting for one of two things: a problem that might float by to which they can attach their solution or a development in the political stream, such as a change of administration that provides a receptive climate for their proposal” (Kingdon 1984: 195).

In a different vein, Baumgartner and Jones’ punctuation equilibrium theory (PET) also emphasizes agenda setting as a non-linear process (Baumgartner and Jones 1993; Jones and Baumgartner 2005). The PET suggests that political processes are most of the time characterized by long periods of stability, or at least only incremental changes. In times of “normal” policy making, things do not change very much. However, sudden large-scale changes do happen, often within a short time span. Hence, PET explains both long time periods with stability – or at least only marginal change – and periods of sudden and significant changes, which are labeled “punctuations”.¹

The punctuations reflect that policy makers process information disproportionately. Following Simon’s (1951) bounded rationality theory, the PET model suggests that issue attention is never proportional to the severity of problems. Most of the time, policy makers

tend to ignore any signal from the environment about a need to change policy. These periods of policy stability are interrupted by dramatic increases in issue attention, in which policy makers tend to overact, paying disproportionate attention to an issue. The arguments behind PET were completely at odds with the incrementalism model of policy change developed by Lindblom (1959) and, later, Wildavsky (1964) for budgetary change, according to which policy makers adopt decisions step by step, following incremental changes from existing policies. Stability is governed by what Baumgartner and Jones (1993) identify as a “policy monopoly”:

Every interest, every group, every policy entrepreneur has a primary interest in establishing a monopoly – a monopoly on political understanding concerning the policy of interest, and an institutional arrangement that reinforces that understanding ... Participation in a policy monopoly is structured by two things: the formal or informal rules of access discourage the participation of outsiders, and the prevalent understanding of the policy are so positive that they evoke only indifference by those not involved (thereby insuring their continued noninvolvement). (Baumgartner and Jones 1993: 7)

The policy monopoly works to maintain an equilibrium around a given policy question and to counterbalance any changes in the principles and values behind existing policies (what they call policy image) coming in from the environment.

Occasionally, however, some issues come to the forefront, and major policy actors begin discussing them. At these times, policies can change very rapidly. Rapid policy change occurs as a result of two main processes: mimicking and attention shifting. The first operates when people observe the behavior of others and act accordingly. The second operates because people, in the words of Herbert Simon, are serial information processors. They attend to only limited parts of the world at any given time. Since one cannot possibly simultaneously be attuned to all elements of the world around one, people use various informational “short-cuts” in order to make reasonable decisions (Jones and Baumgartner 2005). Most people tend to focus on one or just a few dimensions in making their choices. Yet, once new dimensions gain importance in the political debate, people may shift their attention towards these new dimensions of the issue, following a quite unpredictable change in their behavior.

3. METHODOLOGICAL DEVELOPMENTS

Despite the long theoretical history outlined above, it was not until the work of Baumgartner and Jones (1993) that quantitative research on policy agendas began to develop. In *Agendas and Instability in American Politics* (1993), Baumgartner and Jones tracked attention to a number of specific policy questions such as pesticides, smoking, alcohol, and nuclear power. Attention to these policy questions was tracked by keyword search using several sources: the *Readers' Guide to Periodical Literature* and *New York Times* to measure the public agenda, and Congressional hearings to measure the political agenda. What was unique about the data used by Baumgartner and Jones in *Agendas and Instability* was that they allowed for a focus on long-term attention dynamics. Thus, the times series began right after World War II or even earlier.

The datasets that Baumgartner and Jones developed for the 1993 book were the forerunner of the CAP project. There are, however, some fundamental differences between the datasets that Baumgartner and Jones developed and the CAP datasets. First, the CAP project datasets are developed according to a predefined coding scheme, not on keyword searches for specific

policy questions (see Appendix). This is a relevant methodological advance mainly because keyword search is difficult to apply over long time periods when the language used to discuss a policy question can change dramatically.

Second, CAP project datasets cover the policy agenda in a comprehensive manner, describing its entire development rather than only focusing on specific issues or policy questions such as nuclear power or pesticides. This is a major contribution to the study of agenda setting, allowing to go beyond case studies and to trace the evolution of issue attention across time and countries. Finally, CAP datasets provide information not only about the final coding but also about the actual political activities that have been coded. This is a major methodological contribution to the study of agenda setting, which allows researchers to go further in the analysis of important theoretical questions. For example, the CAP coding scheme codes attention only, leaving aside the tone relating to the policy questions. Coding tone for each policy question would be a gigantic task. However, to obtain information about the tone, researchers can use the CAP datasets by identifying the issues that respond to their theoretical goals and then use the detailed information for each agenda item to code the tone. A number of additional points about the CAP coding scheme are worth highlighting.

All coding items (hearings, bills, etc.) are assigned to both a major code and a subtopic code. This structure in reality implies that coding is carried out at the level of the subtopics, and afterwards, any researcher can combine the subtopics in whatever ways suits a particular research question. Thus, the major topics are largely a way of providing an overview of the subtopics. However, each major topic also has a “general code” (00) to cover general questions. An example could be a politician stating that “societies need to focus on education”. Each major topic also has an “other category” (99) to deal with very specific questions, concerning, for instance, the environment, that do not fit a specific subcategory.

One decision made by Baumgartner and Jones (1993) that has stayed with the coding scheme ever since is the decision to code each coding item into one subtopic only. This is not to argue that there are not cases in which a coding item could be coded in multiple subtopics, such as a hearing on energy saving and climate change. The cost for this would, however, be to add considerable additional complexity to the coding: What if a coding item touches upon more than two subtopics, and what if they are not equally present? Thus, coding each item into only one category is a simple coding rule that focuses on the major content of an item rather than capturing all aspects. Further, in case multiple subtopics are equally relevant, items are coded in the “general” category (00) of a major category.

When looking at the CAP scheme, one should be aware that the scheme is supposed to measure the *policy* content of various political agendas. The major topics are classic policy issues such as transportation, the environment, the labor market, and so on. The coding thus tries to capture the policy content or the policy measures included in the items coded. This is logical given that the scheme was developed with policy agenda-setting theory as the guiding theory and a focus on explaining policy change. This focus on policy means or policy content clearly differentiates the CAP scheme from, for instance, the coding scheme used by the Comparative Manifesto Project (CMP) (cf. Green-Pedersen 2019). The CMP coding scheme was developed to capture party ideology in order to place political parties on a left–right scale. Focus in the coding was thus on the ideological aims of a given item rather than the policy content.

From the outset, the structure of the CAP coding scheme has been 21 major topics and around 225 subtopics. A relevant question is what the underlying idea of this structure is. For

instance, why have 21 major topics as the organizational structure rather than 18 or 20? There is no definite answer to this. However, the idea of major topics is close to the idea of “policy issues”, which play an important role in the organization of politics in most countries. Standing committees in legislatures are typically organized around policy issues such as education, labor market, health care, or the environment. Ministerial portfolios also resemble policy issues such as the minister of education, agriculture, or defense. Political parties also typically have spokespersons for particular policy issues. Thus, policy issues are an organizational structure of politics – it is not an invention of the CAP coding scheme but a phenomenon of politics in the real world.

The subtopics of the CAP coding scheme mostly resemble what can be labeled “policy questions”. These are the type of questions that Baumgartner and Jones focused on in their 1993 book concerning, for instance, nuclear power and tobacco. They resemble what politicians deal with when implementing policy reforms or striking larger policy deals. Politicians rarely make policy reforms that cover an entire policy issue, such as education. They typically focus on a specific policy question such as primary schools or universities.

To sum up, the exact number of policy issues and their content are not something that can be deduced theoretically, but are a pragmatic consideration of how many and which categories are useful in terms of describing the policy content of an agenda. However, the idea of approaching politics through major topics and subtopics resembles how real-world politics is organized.

The CAP coding scheme was developed for a US context, but within the last decades, similar datasets have been developed for a large number of countries, in particular in Western Europe, but also countries such as Australia, Brazil, Canada, Hong Kong, Israel, New Zealand, Russia, South Korea, and Turkey (see Baumgartner, Breunig, and Grossman 2019). This shows that the CAP methodology and the PET framework are applicable also outside liberal democracies as long as the political systems disclose their activities in parliament, such as executive orders, debates, and speeches. The fact that new countries are constantly added to the CAP community is a testimony to its success as a framework to analyze public policy. CAP data have further been developed for the American states of Florida and Pennsylvania, the European Union, and the 98 Danish municipalities. The development of CAP data for all these countries has implied coding of data sources that were not part of the original US project. This includes sources such as party manifestos, coalition agreements, cabinet agenda meetings, local council meeting agendas, and parliamentary control instruments such as questions to the minister and various forms of motions and interpellations. A number of data sources are also similar to the ones coded for the US, such as speeches by the head of governments, bills, high court decisions, media data (typically newspaper articles), and, more recently, social media.

The expansion of the CAP data to such a broad range of countries and also to new data sources has been a tough test for the coding scheme originally developed for the US federal level and, especially, the US Congress. The guiding principle for all the countries has been to use the original CAP coding scheme as the starting point and, then, to carry out necessary national adaptations to secure that the coding scheme captured important national trends in policy agendas. Such adjustments could be of different kinds with different implications for the cross-national comparability of the data. When new subtopics are developed by splitting particular subtopics into two or more new subtopics, this is unproblematic. The subtopics can just be merged again when comparing with other countries. What is more problematic is when new subtopics are created with a content belonging to different subtopics across different

Table 6.1 Major topics, Comparative Agendas Project coding system

Code	Description	Code	Description
1	Macroeconomic	14	Housing
2	Civil Rights	15	Domestic Commerce
3	Health	16	Defense
4	Agriculture	17	Technology
5	Labor	18	Foreign Trade
6	Education	19	International Affairs
7	Environment	20	Government Operations
8	Energy	21	Public Lands
9	Immigration	23	Culture
10	Transportation		
12	Law and Crime		
13	Social Welfare		

major categories. In that case, the new subtopic cannot just be merged into a subtopic in the original CAP coding scheme.

None of the countries have carried out fundamental revisions to the original CAP coding scheme when applying it, but at the same time, the changes made were in some cases more than marginal. This raised an issue of cross-national comparability of the data. Given that each country has made its own version of the codebook, a cross-national realignment of the datasets was necessary. This process was organized by Shaun Bevan (see Bevan 2018) and led to a new version of the CAP coding scheme, known as the CAP master codebook (see <http://www.sbevan.com/cap-master-codebook.html>). This codebook has 21 major topics and 213 subtopics (see Table 6.1). The two new major topics added are immigration and cultural policy. Compared to the original US codebook, the master codebook reflects two types of changes. One is adjustments due to the application of the original CAP scheme in a much wider comparative context. Cultural policy might be a minor policy question in the US Congress but an important policy issue in many countries, including ministries and parliamentary standing committees for cultural policy. The other is adjustments due to the emergence of new policy questions that also apply to the US. For instance, a subtopic was added to the master codebook to capture domestic responses to terrorism.

The aim of the master codebook was to have a codebook that would “bridge” all the national datasets. Thus, the national datasets can be recoded to match the master codebook. This goal has not been accomplished in every detail, but the result comes close, and through the development of the master codebook, the CAP data make it possible to track policy agendas over long time periods and in great detail in the countries covered and also to compare with a wide range of other countries.

4. POLITICAL AGENDAS

In general, we refer to the political agenda as the list of issues to which political institutions are paying some serious attention at any given time (Dearing and Rogers 1996). However, there is not a unique definition of the political agenda. Cobb and Elder made the distinction between the systemic agenda, defined as a general set of political controversies that will be viewed as

falling within the range of legitimate concerns meriting the attention of the polity, and the institutional agenda, defined as a set of concrete items scheduled for active and serious consideration by a particular institutional decision-making body (Cobb and Elder 1983). Kingdon (1984) distinguishes between the governmental agenda, defined as the list of issues a government is paying attention to, and the decision agenda, which includes the issues a government is ready to take a formal decision upon (see Zahariadis 2016 for a review).

The CAP follows a broader definition of the political agenda, as the list of issues that actors in and around the political system pay attention to over time. This includes actors in a wide variety of political institutions such as parliaments, governments, public administrations, and judiciaries. It also includes non-governmental actors such as interest organizations which could be non-governmental organizations, business groups, professional organizations, trade unions, and think tanks. These operate across different levels of governance – supranational, national, and sub-national. Prime ministers, members of parliament, political parties, and/or interest groups have their own agenda and pay attention to issues in a different fashion, taking into account flows of information, policy preferences, and the limits and opportunities of the institutional setting in which they operate. Further, this focus on particular agendas makes it possible to study the interaction between different agendas, as several authors do (Alexandrova 2017; Chaqués Bonafont, Palau, and Baumgartner 2015; Green-Pedersen and Mortensen 2010; Green-Pedersen and Stubager 2010; Jennings and John 2009, 2011; Vliegthart et al. 2016; Walgrave, Soroka, and Nuytemans 2008).

However, agendas also differ because of differences in the actors generating the agenda and their interaction. For instance, Green-Pedersen and Mortensen (2010) have introduced the concept of a “party system agenda” to understand how parties compete to decide which issues they should pay attention to. Their argument is that at any point in time, a handful of issues are the subject of political parties’ attention, and no party can ignore these issues. At the same time, parties constantly compete to influence the issues of tomorrow. Therefore, what characterizes the party system agenda is that it is generated by a limited number of actors. Party systems typically consist of five to ten parties, which are at the same time competitors. Political parties compete for votes, office, and policy influence and behave strategically towards each other. The media or news agenda is also generated by a relatively limited number of news media that interact with each other. Although they are also competitors for readers/viewers, their interactions are also driven by journalistic norms about news value and critical journalism (McCombs 2004: 98–102). One can speak of the media agenda as the agenda of the “news institution” (Cook 1998).

The public agenda is a quite different agenda. The public, that is, the voters, are not a limited group of strategic actors, but a large number of individuals for whom politics is mostly a marginal aspect of their everyday life. The public agenda is typically measured through an MIP (“Most Important Problem”) question (Jennings and John 2009, 2011). Most voters will not spend much time on thinking about this question unless it occurs in a survey, and they are unlikely to act strategically to influence the answer given by other members of the public. In other words, the public agenda emerges when the public is asked about it, not because of the continuous interaction of professional and strategic actors such as political parties and news media. This difference between the public agenda and other agendas has implications for how one theorizes its relation to other agendas. While political parties might strategically try to influence both the party system agenda and other agendas, the public agenda is primarily

responsive to other agendas (Soroka 2002). How can one imagine the public becoming concerned with an issue without this happening through news stories?

5. MEASURES OF THE AGENDA

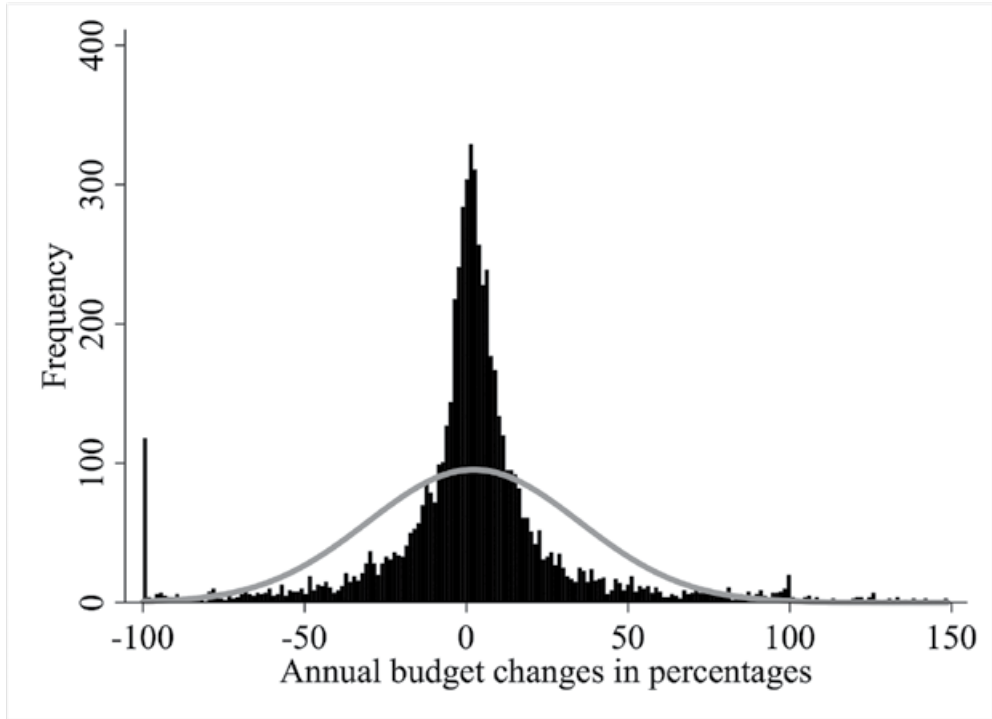
The starting point in describing and analyzing policy agendas is that attention is scarce. The agenda may grow or contract somewhat in the long run in terms of the number of items or subtopics on the agenda, but in the short to medium term, the question is not how large the agenda is but which issue is on the agenda at the expense of other issues (Baumgartner, Jones, and Wilkerson 2011). Therefore, the attention to an issue is usually measured in relative terms (in percentages) instead of absolute terms, and most analyses are focused on developments in the allocation of attention across issues over time (see, e.g. Baumgartner and Jones 1993; Green-Pedersen and Walgrave 2014; John et al. 2013). A number of measures have been developed and applied to describe and analyze variation in the composition of the political agenda.

5.1 Measure of Policy Change – Leptokurtosis

To describe to what extent an agenda is characterized by longer periods of stasis and shorter periods of bursts – what is often referred to as punctuated equilibrium – a measure of leptokurtosis, “the LK score”, has been introduced by Baumgartner and Jones (Baumgartner et al. 2009b; Baumgartner, Jones, and Mortensen 2017). It has become a key measure in much subsequent research (Breunig and Jones 2011; Walgrave and Nuytemans 2009).

Leptokurtosis describes the distribution of changes to the agenda across issues in a given period. Say, attention to education increases by 2%, attention to health drops by 7%, and attention to the military rises by 22%. These percentage changes are summarized in a simple frequency distribution. Leptokurtosis calculates the shape of the distribution and more specifically indicates how far the distribution deviates from a normal distribution in terms of its tails, its peak, and its shoulders (in between the peak and each tail). Quite narrow tails usually characterize a normal distribution whereas a kurtosis distribution has fat tails (huge changes), a taller and narrower central peak (many small changes), and weaker shoulders (few moderate changes). In other words, changes to attention in a kurtosis distribution are either minor in the tall center of the distribution and there are a lot of them – the stasis or long periods of equilibrium – or the changes are rare but huge in the fat tails – the episodic bursts or punctuations. Figure 6.1 graphs the difference between a normal distribution – the black bell-shaped line – and a leptokurtotic distribution marked by the bars.

A normal distribution has a leptokurtosis score of 0.12, and Baumgartner and Jones show that every policy output series in the US, such as hearings, bills, statues, laws, and budgets are non-normally distributed with high leptokurtosis scores (Jones and Baumgartner 2005: 171–83). Baumgartner et al. (2009b) extend this finding to European countries and show that the distribution of changes to attention across issues in various political activities across these countries ranges from .23 to .64, that is, highly leptokurtotic or non-normally distributed. As one moves further into the policy process, the distribution of attention deviates more, and the distribution is more leptokurtotic the more friction the political institutions exhibit. Importantly, Baumgartner et al. (2009b) report that all distributions in the US, Belgium, and



Source: US Policy Agendas Project data.

Figure 6.1 Pooled frequency distribution of annual percentage changes in the US Congressional Budget Authority, FY 1947–2017, in constant 2009 million dollars

Denmark for attention changes in political activities are leptokurtotic. Based on these findings, the authors conclude that PET applies broadly across political systems.

5.2 Measures of Agenda Size – Contraction or Expansion?

To measure if the agenda is contracting or expanding over time, measures of capacity and complexity are relevant (e.g. McCombs and Zhu 1995). Capacity measures the number of items on the agenda, regardless of the number of subtopics in which these items have been categorized. Complexity is concerned with the number of subtopics on the agenda at any given point in time. The more subtopics, the more complexity. Hence, complexity is measured as the number of subtopics from the CAP codebook that are used to code the agenda. As an example, a local council agenda concerned with 18 subtopics might include 27 items, which the council will be discussing on a meeting, that is, about 1.5 items per subtopic. This implies an agenda with a high complexity, but a low capacity in contrast, to say, an agenda with only 13 subtopics but 52 items. Capacity can also be measured simply as the length of, for instance, party manifestos or minutes from parliamentary debates.

Several authors (e.g. Baumgartner and Jones 2015; Green-Pedersen 2007) use these capacity and complexity measures to characterize the development of the political agendas since World War II in Western countries. Across countries, Green-Pedersen (2007) finds that, generally, the agenda has gradually expanded in terms of its capacity and complexity. A main part of this expansion is that new issues such as the environment, crime, and immigration increasingly receive attention in addition to more classic issues such as social policy and the economy. Using the same topics codes, Baumgartner and Jones (2015: chapter 6) identify a “great issue expansion” in the US in the 1960s–1970s, peaking in the 1980s, followed by a sizeable contraction in the 1990s and 2000s. In the 1950s, Congress attended to about 100 different topics annually. At its apex in the mid-1980s, this was closer to 200 topics, that is, almost the entire CAP codebook. In 2010, the Congress was back at its 1950s level. Hence, the US federal government first grew considerably more complex by covering still more policy areas and then reacted to that acceleration.

5.3 Measures of Fragmentation – Entropy

Some agendas are concentrated on a few issues, and some are more spread out. The entropy measure, developed in communication research, reveals this diversity of the agenda. The entropy score is used in an increasing number of studies to understand the allocation of attention (Baekgaard, Mortensen, and Seeberg 2018; Baumgartner and Jones 2015; Boydston 2013; Jennings et al. 2011). There are several ways to calculate the entropy score, and Boydston, Bevan, and Thomas (2014) argue that the Shannon’s H is most appropriate. The Shannon’s H is calculated by multiplying the proportion of the agenda that each item receives by the natural log of that proportion, then taking the negative sum of those products: $-i = \sum p(x_i) \cdot \ln p(x_i)$, where x_i represents an item, $p(x_i)$ is the proportion of the total attention the item receives, and $\ln(x_i)$ is the natural log of the proportion of attention the item receives. If one issue receives all the attention, the Shannon’s H takes the value zero, and the score increases as the spread of attention across all topics becomes more equal (Boydston, Bevan, and Thomas 2014).

An important point in relation to the entropy measure: the score does not change if two issues change position on the agenda. If macroeconomic policy receives 50% of the attention last year and only 25% this year and the opposite happens for crime, then the overall distribution of attention is unchanged and therefore also the entropy score. Hence, the entropy score is useful to characterize the agenda, but it also has a black spot (which the volatility measure below covers).

In a recent book on *The Politics of Information*, Baumgartner and Jones (2015: chapter 5) look at the issue-based structure of the US committee system in order to understand how the US Congress processes information. The authors are interested in the extent to which one committee has monopoly on an issue (a low “issue entropy”) or if committee jurisdictions are overlapping, hence allowing hearings on the same issue in many different committees (a high “issue entropy”) and, conversely, if a committee is focused on a few issues (a low “committee entropy”) or involved in many issues (a high “committee entropy”). Looking over time, Baumgartner and Jones find that entropy, that is, diversity, increased considerably in the 1960s–1980s, so that committees became involved in still more issues, and issues were covered in a more diverse set of committees. In the 2010s, this complexity tapered off somewhat.

Table 6.2 *Hypothetical example of agenda stability*

	Issues (%)			Total (%)
	A	B	C	
Political agenda at t1	40	15	45	100
Political agenda at t2	50	20	30	100

5.4 Measure of Volatility and Issue Overlap

To describe how stable the agenda is, the volatility measure indicates the extent to which the agenda changes its issue content between two periods. The volatility measure can also be used to assess the overlap of two agendas at the same point in time, that is, the agenda of two political parties. Then, the measure is a measure of issue overlap for which it was originally developed by Sigelman and Buell (2004). To understand the logic of this measure, consider a case with two successive executive speeches and three potential issues to address and in which attention to the three issues is distributed as in Table 6.2 (see Mortensen et al. 2011).

In this case, the absolute differences between the two government speeches in terms of issue attention would sum up to 30 ($|40 - 50| + |15 - 20| + |45 - 30|$). Since a value of zero represents perfect agenda stability and a value of 200 represents perfect agenda instability, this hypothetical example would be a case of relatively strong agenda stability between t1 and t2. Standardizing this measure to range between zero and 100 and subtracting from 100 to convert it into a measure of stability rather than instability, the agenda stability (AS) measure can be expressed as

$$Agenda\ Stability_t = 100 - \frac{(\sum_{i=1}^n |GS_t - GS_{t-1}|)}{2}$$

where GS_t and GS_{t-1} is the percentage of the total government speech devoted to a particular issue at Time t and Time $t-1$, and the absolute differences between them are summed over all n of the potential issues on the agendas. Hence, if AS_t equals 100, the issue composition of the government agenda in year t is identical to the issue composition of the government agenda in year $t-1$. On the other hand, if AS_t equals zero, the two successive government agendas have been focused on entirely different issues. A score of, say, 70 for a given year would indicate a 70% overlap between that year’s government agenda and the previous year’s agenda.

Mortensen et al. (2011) have used the volatility measure to understand to what extent the executive agendas in Denmark, the Netherlands, and Britain from the 1950s until recently change when a new party enters office. They find a large degree of overlap. About three-fourths of the issue content of the agenda does not change despite a change to the incumbents. Hence, regardless of the party of the executive, the same issues need attention. Chaqués Bonafont, Palau, and Baumgartner (2015) arrive at a similar conclusion for the case of Spain, and the same occurs for the case of the UK (John et al. 2013).

6. FUTURE DIRECTION OF POLICY AGENDA RESEARCH

Research on the policy agenda has developed tremendously in recent years because of collaborative work and the open source data policy of CAP allowing scholars from diverse backgrounds to use the data. The collection and categorization of millions of agenda items are a major achievement of the CAP community. Yet, one may rightly ask, why be so fascinated with data? The idea is that with high-quality data, which allow comparisons across issues and countries over time, it is possible to test general propositions about the functioning of democracy. Agenda scarcity and information processing is a basic feature of politics, and this means that agenda-setting data are used in a wide range of research agendas. The agenda-setting approach is applied to analyze a wide variety of political actors and institutions. Examples are legislatures (e.g. Adler and Wilkerson 2013; Chaqués Bonafont and Palau 2011), committees (e.g. Jones, Baumgartner, and Talbert 1993), executives (e.g. Mortensen et al. 2011; Seeberg 2017), government departments (Mortensen and Green-Pedersen 2015), the bureaucracy (e.g. Baekgaard, Mortensen, and Seeberg 2018; Workman 2015), parties (e.g. Green-Pedersen and Mortensen 2010; Thesen 2013), interest groups (e.g. Baumgartner et al. 2009a; Chaqués Bonafont and Muñoz Márquez 2016), the media (e.g. Baumgartner and Chaqués Bonafont 2015; Vliegenthart et al. 2016; Walgrave, Soroka, and Nuytemans 2008), and voters (e.g. Jennings and John 2009).

This proliferation of agenda-based research not only brings new insights on how the political agenda is formed but also provides new perspectives on the actors and institutions in and around the political system. Moreover, the data allow more encompassing analysis of a political system, including Spain (Chaqués Bonafont, Palau, and Baumgartner 2015), the UK (John et al. 2013), the US (Baumgartner and Jones 2015), the EU (Alexandrova, Rasmussen, and Toshkov 2016; Alexandrova and Timmermans 2013), and other countries (Green-Pedersen and Walgrave 2014: 201). A key part of the research agenda using policy agendas data also focuses on one particular issue over time, possibly also across countries, including the death penalty (Baumgartner, De Boef, and Boydston 2008), morality issues (Engeli, Green-Pedersen, and Larsen 2012), health care (Green-Pedersen and Wilkerson 2006), pharmaceuticals and food safety (Chaqués Bonafont and Palau 2009), immigration (Green-Pedersen and Krogstrup 2008), euthanasia (Green-Pedersen 2007), crime (Seeberg 2013), and the environment (Seeberg 2016). Hence, the CAP datasets allow researchers to go further in the analysis of research questions that have traditionally been underdeveloped. We conclude by pointing to some of the most interesting research questions for future agenda-setting research.

6.1 Theorizing Issue Characteristics

The CAP datasets allow to identify which issues reach the political agenda. However, we know little about the importance of issue characteristics (Jones and Baumgartner 2005). A few sets of studies have already shown that some issues are just more prone to make it to the political agenda (e.g. Jennings et al. 2011), but so far, it has proven difficult to disentangle issue characteristics in a systematic way. Whereas most recent comparative agenda-setting research aims to show that the agenda-setting theory is fundamental to the political system and therefore applies on diverse issues and across countries (e.g. Baumgartner et al. 2009b), some studies touch on the role of issue idiosyncrasies (see Green-Pedersen and Wilkerson 2006; Soroka 2002).

6.2 Studying Non-Decisions

The CAP datasets are also an important starting point to study non-decisions. From its origins, one of the key motivations to study political agendas is that some issues do not make it to the agenda because some actors “block the door”. This is detrimental because such agenda power means that a policy most likely will not change. However, despite half a century’s research, since the pioneering work of Bachrach and Baratz (1962), systematically studying non-decisions still proves difficult. The challenge for future research is to come up with a credible counterfactual that an issue that did not reach the agenda would have made it to the agenda had the right conditions been present.

6.3 Coding Tone of the Attention

The CAP has collected and topic coded a gigantic number of political activities in recent years. This data collection reveals which issues are on the agenda. The next step is to know how these issues are debated, for instance, whether agriculture is debated as an important export industry or as a source of water pollution. In this vein, an increasing body of research on framing (Baumgartner, De Boef, and Boydston 2008; Boin, ‘t Hart, and McConnell 2009; Rose and Baumgartner 2013) and problem definition (Euchner et al. 2013; Fifer and Orr 2013; Huff and Kertzer 2017) shows that the content of the issue attention decides if attention spurs legislative change and which type of change. Moreover, in their 1993 book, Baumgartner and Jones coded the tone on each issue, that is, whether news items and political activities were positive or negative to the industry, and showed how the balance between positive and negative had to change before the legislation could change. Although it is a major task to systematically code the frame of each political activity at a large scale across countries and back in time, such frame coding of the issues in the existing data has great analytical potential. The challenge is that it already requires quite some local knowledge to just topic code a political activity. To identify the framing requires even more knowledge. However, in terms of understanding attention dynamics, it would be a quantum leap to code this systematically.

6.4 Strengthening the Link between Agenda Setting and Actual Policy Decision

The motivation to investigate agenda setting is that it matters for policy decisions. Decision-makers have to attend to an issue in order to legislate. However, research shows that far-reaching legislation is sometimes enacted quietly or without much prior debate (Culpepper 2010; Hacker 2004). Meanwhile, some issues receive considerable attention, such as the gun shootings in the US without legislative change. As agenda-setting research is often more concerned with the causes of attention rather than the consequences, the link between attention and legislation or budgets offers a fertile ground for future research.

Summing up, taking the next step in agenda-setting research essentially involves taking a step back to carry out more research on the basics of agenda-setting dynamics, namely, non-decisions, issue characteristics, and when attention begets policy change.

NOTE

1. To describe this pattern, Baumgartner and Jones borrowed a term from evolutionary biology: “punctuated equilibrium”, originally coined by Eldredge and Gould (1972) to explain relatively rapid changes in biological species.

REFERENCES

- Adler, S. and Wilkerson, J. (2013). *Congress and the Politics of Problem Solving*. New York: Cambridge University Press.
- Alexandrova, P. (2017). Institutional issue proclivity in the EU: the European Council vs the Commission. *Journal of European Public Policy* 24 (5): 755–74.
- Alexandrova, P., Rasmussen, A., and Toshkov, D. (2016). Agenda responsiveness in the European Council: public priorities, policy problems and political attention. *West European Politics* 39 (4): 605–27.
- Alexandrova, P. and Timmermans, A. (2013). National interest versus the common good: the Presidency in European Council agenda setting. *European Journal of Political Research* 52 (3): 316–38.
- Bachrach, P. and Baratz, M. S. (1962). Two faces of power. *American Political Science Review* 56 (4): 947–52.
- Baekgaard, M., Mortensen, P. B., and Seeberg, H. B. (2018). The bureaucracy and the policy agenda. *Journal of Public Administration Research and Theory* 28 (2): 239–53.
- Baumgartner, F. R., Berry, J. M., Hojnacki, M., Leech, B. L., and Kimball, D. C. (2009a). *Lobbying and Policy Change: Who Wins, Who Loses, and Why*. Chicago, IL: University of Chicago Press.
- Baumgartner, F. R., Breunig, C., Green-Pedersen, C., Jones, B. D., Mortensen, P. B., Nuytemans, M., and Walgrave, S. (2009b). Punctuated equilibrium in comparative perspective. *American Journal of Political Science* 53 (3): 603–20.
- Baumgartner, F. R., Breunig, C., and Grossman, E. (Eds.) (2019). *Comparative Policy Agendas: Theory, Tools, Data*. Oxford: Oxford University Press.
- Baumgartner, F. R. and Chaqués Bonafont, L. (2015). All news is bad news: newspaper coverage of political parties in Spain. *Political Communication* 32 (2): 268–91.
- Baumgartner, F. R., De Boef, S. L., and Boydston, A. E. (2008). *The Decline of the Death Penalty and the Discovery of Innocence*. New York: Cambridge University Press.
- Baumgartner, F. R. and Jones, B. D. (1993). *Agendas and Instability in American Politics*. Chicago, IL: University of Chicago Press.
- Baumgartner, F. R. and Jones, B. D. (2015). *The Politics of Information*. Chicago, IL: University of Chicago Press.
- Baumgartner, F. R., Jones, B. D., and Mortensen, P. B. (2017). Punctuated equilibrium theory: explaining stability and change in policy making. In C. M. Weible and P. A. Sabatier (Eds.), *Theories of the Policy Process* (4th edition). Boulder, CO: Westview Press, 55–102.
- Baumgartner, F. R., Jones, B. D., and Wilkerson, J. (2011). Comparative studies of policy dynamics. *Comparative Political Studies* 44 (8): 947–72.
- Bevan, S. (2019). “Gone Fishing: The Creation of the Comparative Agendas Project Master Codebook”. pp. 17–35. In *Comparative Policy Agendas: Theory, Tools, Data*, edited by Frank Baumgartner, Christian Breunig, and Emiliano Grossman. Oxford: Oxford University Press.
- Boin, A., ‘t Hart, P., and McConnell, A. (2009). Crisis exploitation: political and policy impacts of framing contests. *Journal of European Public Policy* 16 (1): 81–106.
- Boydston, A. E. (2013). *Making the News: Politics, the Media, and Agenda Setting*. Chicago, IL: University of Chicago Press.
- Boydston, A. E., Bevan, S., and Thomas, H. F. (2014). The importance of attention diversity and how to measure it. *Policy Studies Journal* 42 (2): 173–96.
- Breunig, C. and Jones, B. D. (2011). Stochastic process methods with an application to budgetary data. *Political Analysis* 19 (1): 103–17.

- Chaqués Bonafont, L. and Muñoz Márquez, L. M. (2016). Explaining interest group access to parliamentary committees. *West European Politics* 39 (6): 1276–98.
- Chaqués Bonafont, L. and Palau, A. M. (2009). The dynamics of policy change: a comparative analysis of the food safety and pharmaceutical policy in Spain. *Journal of Public Policy* 29 (1): 103–26.
- Chaqués Bonafont, L. and Palau, A. M. (2011). Assessing the responsiveness of Spanish policymakers to the priorities of their citizens. *West European Politics* 34 (4): 706–30.
- Chaqués Bonafont, L., Palau, A. M., and Baumgartner, F. R. (2015). *Agenda Dynamics in Spain*. Basingstoke: Palgrave Macmillan.
- Cobb, R. and Elder, C. (1971). The politics of agenda-building: an alternative perspective for modern democratic theory. *Journal of Politics* 33 (4): 892–915.
- Cobb, R. and Elder, C. (1983). *Participation in American Politics: The Dynamics of Agenda-Building*. Baltimore, MD: Johns Hopkins University Press.
- Cook, T. (1998). *Governing with the News*. Chicago, IL: University of Chicago Press.
- Culpepper, P. D. (2010). *Quiet Politics and Business Power: Corporate Control in Europe and Japan*. New York: Cambridge University Press.
- Dearing, J. W. and Rogers, E. M. (1996). *Agenda-Setting*. Thousand Oaks, CA: Sage.
- Eldredge, N. and Gould, S. (1972). Punctuated equilibria: an alternative to phyletic gradualism. In F. Ayala and J. Avise (Eds.), *Essential Readings in Evolutionary Biology*. Baltimore, MD: Johns Hopkins University Press, 239–72.
- Engeli, I., Green-Pedersen, C., and Larsen, L. (Eds.) (2012). *Morality Politics in Western Europe*. Basingstoke: Palgrave Macmillan.
- Euchner, E.-M., Heichel, S., Nebel, K., and Raschzok, A. (2013). From “morality” policy to “normal” policy: framing of drug consumption and gambling in Germany and the Netherlands and their regulatory consequences. *Journal of European Public Policy* 20 (3): 372–89.
- Fifer, N. and Orr, S. K. (2013). The influence of problem definitions on environmental policy change: a comparative study of the Yellowstone wildfires. *Policy Studies Journal* 41 (4): 636–53.
- Green-Pedersen, C. (2007). The conflict of conflicts in comparative perspective: euthanasia as a political issue in Denmark, Belgium, and the Netherlands. *Comparative Politics* 39 (3): 273–91.
- Green-Pedersen, C. (2019). Issue attention in West European party politics: CAP and CMP coding compared. In F. R. Baumgartner, C. Breunig, and E. Grossman (Eds.), *Comparative Policy Agendas: Theory, Tools, Data*. Oxford: Oxford University Press, 373–90.
- Green-Pedersen, C. and Krogstrup, J. (2008). Immigration as a political issue in Denmark and Sweden. *European Journal of Political Research* 47: 610–634.
- Green-Pedersen, C. and Mortensen, P. (2010). Who sets the agenda and who responds to it in the Danish Parliament? *European Journal of Political Research* 49 (2): 257–81.
- Green-Pedersen, C. and Stubager, R. (2010). The political conditionality of mass media influence: when do parties follow mass media attention? *British Journal of Political Science* 40 (3): 663–77.
- Green-Pedersen, C. and Walgrave, S. (2014). *Agenda Setting, Policies, and Political Systems: A Comparative Approach*. Chicago, IL: University of Chicago Press.
- Green-Pedersen, C. and Wilkerson, J. (2006). How agenda-setting attributes shape politics: basic dilemmas, problem attention and health politics developments in Denmark and the US. *Journal of European Public Policy* 13 (7): 1039–52.
- Hacker, J. S. (2004). Privatizing risks without privatizing the welfare state: the hidden politics of social policy retrenchment in the United States. *American Political Science Review* 98 (2): 243–60.
- Huff, C. and Kertzer, J. D. (2017). How the public defines terrorism. *American Journal of Political Science* 62 (1): 55–71.
- Jennings, W., Bevan, S., Timmermans, A., Breeman, G., Brouard, S., Chaqués Bonafont, L., Green-Pedersen, C., John, P., Mortensen, P. B., and Palau, A. M. (2011). Effects of the core functions of government on the diversity of executive agendas. *Comparative Political Studies* 44: 1001–30.
- Jennings, W. and John, P. (2009). The dynamics of political attention: public opinion and the Queen’s Speech in the United Kingdom. *American Journal of Political Science* 53 (4): 838–54.
- Jennings, W. and John, P. (2011). Punctuations and turning points in British politics: the policy agenda of the Queen’s Speech, 1940–2005. *British Journal of Political Science* 40 (3): 561–86.
- John, P., Bertelli, A., Jennings, W., and Bevan, S. (2013). *Policy Agendas in British Politics*. Basingstoke: Palgrave Macmillan.

- Jones, B. D. and Baumgartner, F. R. (2005). *The Politics of Attention: How Government Prioritizes Problems*. Chicago, IL: University of Chicago Press.
- Jones, B. D., Baumgartner, F. R., and Talbert, J. (1993). The destruction of issue monopolies in Congress. *American Political Science Review* 87 (3): 657–71.
- Kingdon, J. (1984). *Agendas, Alternatives, and Public Policies*. New York: HarperCollins.
- Lindblom, C. (1959). The science of muddling through. *Public Administration Review* 19 (2): 79–88.
- McCombs, M. (2004). *Setting the Agenda: The News Media and Public Opinion*. Cambridge: Polity Press.
- McCombs, M. and Zhu, J.-H. (1995). Capacity, diversity, and volatility of the public agenda. *Public Opinion Quarterly* 59 (4): 495–535.
- Mortensen, P. B. and Green-Pedersen, C. (2015). Institutional effects of changes in political attention: explaining organizational changes in the top bureaucracy. *Journal of Public Administration Research and Theory* 25 (1): 165–89.
- Mortensen, P. B., Green-Pedersen, C., Breeman, G., Chaqués Bonafont, L., Jennings, W., John, P., Palau, A. M., and Timmermans, A. (2011). Comparing government agendas: executive speeches in the Netherlands, United Kingdom, and Denmark. *Comparative Political Studies* 44 (8): 973–1000.
- Rose, M. and Baumgartner, F. R. (2013). Framing the poor: media coverage and U.S. poverty policy, 1960–2008. *Policy Studies Journal* 41 (1): 22–53.
- Schattschneider, E. (1960). *The Semisovereign People: A Realist's View of Democracy in America*. New York: Holt, Rinehart & Winston.
- Seeberg, H. (2013). Parties and policies: on the opposition's influence on policy through issue politicization. *Journal of Public Policy* 33 (1): 89–117.
- Seeberg, H. (2016). Opposition policy influence through agenda-setting: the environment in Denmark, 1993–2009. *Scandinavian Political Studies* 39 (2): 185–206.
- Seeberg, H. (2017). What can a government do? Government issue ownership and real-world problems. *European Journal of Political Research* 56 (2): 346–63.
- Sigelman, L. and Buell, E. (2004). Avoidance or engagement? Issue convergence in US presidential campaigns 1960–2000. *American Journal of Political Science* 48 (4): 650–661.
- Simon, H. A. (1951). A formal theory of the employment relationship. *Econometrica* 19 (3): 293–305.
- Soroka, S. (2002). *Agenda-Setting Dynamics in Canada*. Vancouver: UBC Press.
- Thesen, G. (2013). When good news is scarce and bad news is good: government responsibilities and opposition possibilities in political agenda-setting. *European Journal of Political Research* 52: 364–89.
- Vliegenthart, R., Walgrave, S., Baumgartner, F., Bevan, S., Breunig, C., Brouard, S., Chaqués Bonafont, L., Grossman, E., Jennings, W., Mortensen, P. B., Palau, A., Sciarini, P., and Tresch, A. (2016). Do the media set the parliamentary agenda? A comparative study in seven countries. *European Journal of Political Research* 55 (2): 283–301.
- Walgrave, S. and Nuytemans, M. (2009). Friction and party manifesto change in 25 countries, 1945–98. *American Journal of Political Science* 53 (1): 190–206.
- Walgrave, S., Soroka, S., and Nuytemans, M. (2008). The mass media's political agenda-setting power. *Comparative Political Studies* 41: 814–36.
- Wildavsky, A. (1964). *The Politics of the Budgetary Process*. Boston, MA: Little, Brown.
- Workman, S. (2015). *The Dynamics of Bureaucracy in the US Government*. Cambridge: Cambridge University Press.
- Zahariadis, N. (Ed.) (2016). *Handbook of Public Policy Agenda-Setting*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.

APPENDIX

List of Subtopics

1. Macroeconomics

- 100. General
- 101. Inflation and interest rates
- 103. Unemployment
- 104. Monetary policy, the Bank of Spain, monetary reserve discount rate
- 105. Budgets and spending budget law
- 107. Taxes, tax policy and tax reform
- 108. Industrial policy
- 110. Control and stabilization of prices
- 199. Others

2. Civic rights

- 200. General
- 201. Ethnic minorities and racial discrimination
- 202. Gender discrimination and rights
- 204. Age discrimination
- 205. Discrimination of people with illnesses or disabilities
- 206. Voting rights, political participation and representation
- 207. Freedom of speech and religion. Equal rights in general. Abortion
- 208. Right to privacy and access to information
- 209. Activities against the state
- 299. Others

3. Health

- 300. General
- 301. General reforms of the National Health System (NHS)
- 302. General questions about the coverage of the NHS
- 321. Regulation of the pharmaceutical industry, medical devices and clinical labs
- 322. Sanitary facilities, hospitals, construction in the health system
- 323. Agreements between the NHS and private companies
- 324. Medical malpractice, fraud, abuse and compensation systems
- 325. Human resources, education and training. Health manpower
- 331. Disease prevention and health promotion
- 332. Child health
- 333. Mental diseases
- 334. Long-term treatment, rehabilitation services, hospice and aging
- 335. Pharmaceutical expenditure, government consumption and drug prices
- 341. Tobacco
- 342. Alcohol, control of illegal drugs

- 398. Research and development in health
- 399. Others

4. Agriculture

- 400. Agriculture: General
- 401. Agricultural trade
- 402. Subsidies and agricultural regulation
- 403. Food inspection and safety
- 404. Agricultural marketing and promotion
- 405. Animal and crop disease and pest control
- 408. Fishing and hunting policy
- 498. Research on agriculture and livestock
- 499. Other

5. Labor

- 500. General
- 501. Working conditions, work accidents and compensation schemes
- 502. Employment training and workforce development
- 503. Pensions and early retirement. Employee benefits
- 504. Labor unions and employee relations
- 505. Employment policy and collective negotiation
- 506. Youth and child employment
- 529. Labor and immigration
- 599. Others

6. Education

- 600. General
- 601. Higher education
- 602. Elementary and secondary education
- 603. Education of students with difficulties
- 604. Vocational education. Professional training
- 606. Special education for disabled students
- 607. Education excellence
- 698. Research on education
- 699. Others

7. Environment

- 700. General
- 701. Water quality, pollution and conservation of the coast
- 703. Waste disposal
- 704. Hazardous and toxic waste
- 705. Air pollution, global warming and noise pollution

- 707. Recycling
- 708. Indoor environmental hazards
- 709. Species and forest protection
- 711. Land and water conservation
- 798. Research and development
- 799. Others

8. Energy

- 800. General
- 801. Nuclear energy
- 802. Electricity and hydroelectricity
- 803. Natural gas and oil (including offshore oil and gas)
- 805. Coal and mining
- 806. Alternative and renewable energy
- 807. Energy conservation
- 898. Research and development
- 899. Other

9. Immigration

- 900. Immigration and refugee issues

10. Transportation

- 1000. General
- 1001. Mass transportation and safety
- 1002. Highway and roads construction, maintenance, and safety
- 1003. Airports, air traffic control and safety
- 1005. Railroad transportation and safety
- 1007. Maritime issues and naval industry
- 1010. Public works (infrastructure development)
- 1098. Research and development in transportation
- 1099. Other

12. Law and crime

- 1200. General
- 1201. Police and crime fighting authorities
- 1202. Organized crime and financial related crime. Tax fraud. White collar crime
- 1203. Illegal drug production, trafficking, and control
- 1204. Judiciary system and court administration
- 1205. Prisons
- 1206. Youth crime
- 1207. Child abuse and child pornography
- 1208. Domestic violence, gender violence and family issues

- 1210. Criminal and civil code
- 1211. Crime prevention, riots
- 1227. Terrorism and counter-terrorism
- 1299. Other

13. Social welfare

- 1300. General
- 1302. Poverty and assistance for low-income families
- 1303. Elderly issues and elderly assistance programs
- 1304. Assistance to the disabled and handicapped
- 1305. Social services and volunteer associations
- 1308. Work–life balance and child care
- 1399. Other

14. Housing

- 1400. General
- 1401. Housing and community development
- 1403. Urban economic development and general urban issues
- 1404. Housing policy in rural areas
- 1405. Rural economic development and general urban issues
- 1406. Public housing programs
- 1408. Elderly and handicapped housing
- 1409. Housing assistance for homeless and homeless issues
- 1499. Other

15. Domestic commerce

- 1500. General
- 1501. Banking system and financial institution regulation
- 1502. Securities and commodities regulation
- 1504. Mortgages, credit cards and other services of the credit market
- 1505. Insurances regulations
- 1507. Bankruptcy
- 1520. Antitrust regulation and corporate management
- 1521. Small and medium enterprises issues
- 1522. Copyrights and patents
- 1523. Domestic disaster relief (natural disasters and accidents)
- 1524. Tourism
- 1525. Consumer safety, consumer fraud and data protection
- 1526. Sports, lottery and gambling
- 1599. Others

16. Defense

- 1600. General
- 1602. NATO and other defense alliances
- 1603. Military intelligence, CIA, espionage
- 1604. Military readiness and capabilities
- 1605. Arms control and nuclear nonproliferation
- 1606. Military aid and weapons sales to other countries
- 1608. Human resources and military personnel
- 1610. Military procurement and weapons system acquisitions
- 1611. Military installations, properties and building construction
- 1614. Military environmental compliance
- 1615. Civil protection services and armed forces
- 1616. Civilian personnel and employment by the defense industry
- 1617. Defense contracts
- 1619. Direct war related issues
- 1620. Human rights' violations in war. Relief of claims against military forces
- 1698. Research and development on military issues
- 1699. Other

17. Technology

- 1700. General
- 1701. Space missions and research
- 1704. Satellites and other space technology with commercial use
- 1705. Science technology transfer and scientific international cooperation
- 1706. Telecommunication and telephone services
- 1707. Media
- 1708. Weather forecasting and geologic issues
- 1709. Computer industry and computer security
- 1798. Research and development
- 1799. Other

18. Foreign trade

- 1800. General
- 1802. Trade agreements, disputes and regulation
- 1803. Export promotion and regulation
- 1804. Overseas private investment and Spanish investments abroad
- 1806. Productivity and competitiveness. Spain balance of payments
- 1807. Imports and regulation of imports
- 1808. Exchange rates and Forex
- 1899. Other

19. International affairs

- 1900. General
- 1901. Foreign aid
- 1902. International agreements on environmental issues
- 1905. Developing countries issues
- 1906. International finance system and economic development organizations
- 1910. Western Europe and Common Market issues
- 1921. Specific country/region
- 1925. Human rights
- 1926. International organizations and international NGOs
- 1927. International terrorism
- 1929. Diplomacy
- 1999. Others

20. Government operations

- 2000. General
- 2001. Intergovernmental relations and local government
- 2002. Public administration efficacy. Bureaucracy oversight
- 2003. Postal service
- 2004. Civil Service, government employee benefits
- 2005. Appointments and nominations
- 2006. Awards and public honors. Medals
- 2007. Government procurement, contracts and corruption
- 2008. Privatization of public sector. Government property management
- 2009. Central tax administration
- 2011. Government branch relations, Parliament and Constitution
- 2012. Regulation of political activities, elections and election campaigns
- 2015. Claims against the government
- 2030. National holidays
- 2099. Others

21. Public lands

- 2100. General
- 2101. Natural parks and historic sites
- 2103. Natural resources, forest managements and public lands
- 2104. Water and sea resources: development, public works and harbors
- 2199. Others

23. Culture

- 2300. General
- 2301. Cinema, theatre, music and dance
- 2302. Publication of books and literary works in general
- 2399. Others

27. Climate (Only media coding)

2700. General

29. Sports (Only media coding)

2900. General

30. Death notices (Only media coding)

3001. Natural death

3002. Violent death

3099. Others

7. Comparing historical cases: advances in comparative historical research

Grace Jaramillo

1. INTRODUCTION

Comparative historical analysis (CHA) has become a reliable approach for assessing public policy change in small-N and medium-N studies (Barzelay and Gallego 2010; Gallego, Barbieri, and Gonzalez 2017; Mele and Ongaro 2014). The strength of CHA resides precisely in its capacity for tracing all possible causal variables influencing policy change across space and time, a paramount condition for theory-building and theory-testing. CHA combines techniques like process tracing, archival research, statistical methods, and in-depth interviews to assess causal process observations. Arguably, CHA is the most suitable method to understand and explain long-term public policy issues like macroeconomic policy, industrial policy, social security, or the welfare state. Works in this tradition have become foundational pieces to understand the state and its public policy process in the modern era (Hall 1986; Pedersen 1995; Thelen 2003).

The current debate among CHA scholars is, however, how to improve comparisons across time. Paul Pierson (2004) has been one of the most salient defenders of adequate time selection and long-term comparisons across cases. Falletti and Mahoney (2015) advanced a more elaborated description on how not only time horizons but also elapses between critical junctures and intervening events have a profound impact in explaining policy outcomes. Time frames could be short in extreme, misleading the identification of variables; or so extensive as to render the identification of causal mechanisms unattainable. Moreover, the identification of time frames and time horizons in policy analysis rarely pays attention to broader political or economic cycles that can potentially be determinant in assessing causality and agency in the process of policy design or reform. Either for rational institutionalism that uses the concept of punctuated equilibrium or comparative historical analysis, the methodology to define the starting period of analysis is crucial to any attribution of causality. The very issue of determining the strength of events as critical junctures, capable of triggering changes or a wave of reforms, demands a more rigorous selection of time frames that validates not only the correct selection of cycles of change but also the attribution of causality.

This chapter discusses the definition of time frames and time horizons for comparative policy analysis and proposes an index of policy intervention as a suitable tool to validate them. Time frame validation becomes extremely useful in cross-comparisons that entail long-term cycles of policy change. The index helps to organize and understand the bulk of data resulting from process tracing exercises.

The remainder of this chapter is organized in four main parts. The next two sections review the theoretical underpinnings of time in the context of CHA, including the most recent debates regarding time and sequence. Then the index is presented as a key methodological tool. The following part discusses the index applied to a specific small-N case study comparison in

industrial policy. The final part deals with its applicability to other issues and policy cases and cross-comparison exercises, including improving both the determination of policy cycles and identification of causal process mechanisms.

2. COMPARATIVE HISTORICAL ANALYSIS

Historical institutionalism (HI) was initially concerned with how institutions influenced policy outcomes. Over time, HI has become a powerful middle-range theory that understands social processes as the permanent interaction of institutions, context, and agency (Steinmo, Thelen, and Longstreth 1992). Under HI, policy is seen as an institution, since it represents the set of “formal and informal procedures, routines, norms and conventions embedded in the organizational structure of the polity or political economy” (Hall and Taylor 1996: 938).

The origins of HI can be traced back to seminal works of institutional sociology like Barrington Moore’s (1966) *Social Origins of Dictatorship and Democracy* and Theda Skocpol’s (1979) *States and Social Revolutions*. They were some of the first contributions to examine the long-term impacts of institutional configurations of state action in restricting pathways for future action. The lesson for researchers, however, was that the identification of *when* and *how* change happens matters as much as the transformation itself in long-term processes of policy making. In areas like social security, health, education, macroeconomic management and notably industrial policy the way the events unfold, in what order and for how long, can shape the effects of those policies and determine their resilience or not over time. However, the main contribution of HI was certainly its commitment to unpack the study of the state (Evans, Rueschemeyer, and Skocpol 1985). Until then, the state was considered a black box with no discernible autonomy or organizational capacity to influence changes on its own. Over time, HI developed a research agenda and refined research approaches to assess long-term patterns of policy change that delineate the contours of what is now known as the CHA approach.

CHA has built a set of conceptual and methodological tools to study change. The central focus of this research agenda is to explain macro-level aspects of policy and multi-causal mechanisms influencing policy outcomes (Mahoney and Thelen 2015: 4). This ambitious approach has invited also criticism. Radaelli, Dente, and Dossi (2012: 539–40) have pointed out four pitfalls: “institutional determinism”, when only institutional frameworks are considered as direct determinants of change; “drop in the box”, when typologies become straitjackets to policy understanding; “second-best residual explanations”, when institutional variables are mixed and confused with policy-level variables confounding the causal inference process; and lastly, “theoretical conjectures without foundational mechanisms”, when policy outcomes are shoehorned into the institutional approach without properly explaining the causal mechanisms.

To directly address criticism of institutional determinism and drop in the box issues, recent contributions have advanced a rather nuanced understanding of causal mechanisms of change, emphasizing endogenous triggers of change as much as looking at different sources of agency in the process and its intertwining relationship with other variables at the institutional and policy levels (Capoccia 2015, 2016: 1096; Conran and Thelen 2016).

However, CHA has yet to find tools to address issues of second-best residual explanations that sometimes go unnoticed. Notwithstanding, CHA has moved in the last years to refine its assessment of foundational and causal mechanisms of policy change through its handling of

time sensitive methods. In fact, CHA's most important contribution to policy analysis resides in its capacity to identify causal mechanisms and outcomes within specific structures of time (Mahoney and Thelen 2015).

The central reason for refining time and sequence in policy analysis is because it explains the conditions under which change happens. Among the different types of change, path-dependent trajectories and slow-movement cumulative causes and consequences are increasingly important. Path-dependent trajectories tend to be over diagnosed (Pierson 2004) while slow movements tend to be under diagnosed (Mahoney and Thelen 2010). Path dependency is the type of change where a critical juncture triggers a long-term process of change that tends to be self-reinforcing. Every event along the sequence only reinforces the changing trend, inviting positive feedbacks at every turn (Levi 1997: 28). Policies that enter into a path-dependent process are difficult to reverse as reform becomes locked-in over time (Mahoney 2000: 507). A slow-moving process of policy change is less evident and could appear in different cycles and schemes. The central aspect of this process is that it is not self-evident and changes become iterative with no apparent critical juncture triggering them (Mahoney and Thelen 2010: 5).

3. MAXIMIZING PROCESS TRACING RESULTS

It is worth noting that CHA rests on process tracing as the core method to collect causal process observations along the timeline for inquiry. Process tracing is the formal research action of "acquiring more observable implications of a theory from units at the lowest level of aggregation" (George and McKeown 1985: 35). Andrew Bennett (2010: 208) provides a more specific definition of process tracing as the examination of "diagnostic pieces of evidence within a case". The unveiling of intervening causal mechanisms depends on the capacity of the investigator to find and document as many observational facts as possible. Causal process observations are the qualitative method's best option to increase inferential leverage according to Collier, Brady, and Seawright (2010: 507). In other words, process tracing is the best method for both pursuing a scientific model of data collection – even for qualitative studies – and maximizing its capacity to identify causal process observations (CPOs) to gain leverage in the attribution of causality or, in CHA terms, to understand institutional change.

CHA maximizes process tracing results by providing a precise understanding of the strength and leverage of each CPO over determined periods of time. For the researcher, the process of inquiry starts with the definition of time horizons of analysis and within them, three elements that are essential components for the organization of any assessment of causal process observations, namely: events, critical junctures, and sequences.

For CHA, *time horizons* have to be defined with at least two elements in mind: the outcome under study and the initial event that purportedly originated the outcome. However, Pierson (2004) warns against defining extremely short time frames, especially when it comes to public policy. In his words, "failure to recognize the extent to which public policy outcomes are cumulative and slow moving can easily lead social scientists astray" (Pierson 2004: 91). There is a clear advantage in establishing longer time frames for policy analysis, at least long enough to understand the process of policy transition from one paradigm to the next. The challenge is determining the time horizon that renders an adequate account of the outcome of interest. Moreover, this chapter advances the idea that within-case analysis of comparable time frames could become a useful tool to validate attributions of causality but also to assess the extent to

which the time frame under analysis is the correct one and corresponds to the policy change under study.

Following the line of time, there are three concepts that help define an adequate time frame for analysis and more importantly, help to determine the possible trajectory of changes: critical junctures, events, and sequence.

Critical juncture is a key concept of HI in general and CHA in particular. It provides the most important tool to define the origin and strength of institutional reforms. Collier and Berins-Collier (1991: 29) started with a general and overarching definition of critical juncture as “a period of significant change which typically occurs in distinct ways in different countries and which is hypothesized to produce distinct legacies.” A critical juncture is a momentous time, usually provoked by external or domestic shocks, that opens a wave of changes. But how do we determine that the juncture is critical? Pierson’s (2004: 135) response is that “they are critical because they put institutional arrangements on paths or trajectories that are difficult to alter.” Major economic or political shocks, wars, legitimacy crisis, and *coups d’état* are examples of what can constitute critical junctures as departure points for causal process observations about change. However, critical junctures will always be context specific and historically informed. For instance, presidential coups and impeachments may not be critical junctures in countries where they are pervasive. Or they might be once-in-a-lifetime type of cases. Recessions are not per se critical junctures unless they have run out of control and have triggered dramatic consequences that are positively reinforced. Thus, the attribution of critical juncture to a CPO usually is an *ex post* business for the comparative researcher, unless the event itself is too dramatic to even doubt.

A critical juncture is intrinsically linked to mechanisms of increasing returns, dramatic pushes for reform or, at the very least, clear paths of unavoidable change, either because this moment creates positive feedbacks or because it creates a strong process of contestation (Mahoney 2000). The institutional consequences of critical junctures as the cornerstone of change highlight the need of a conscious validation of their strength compared to the rest of the developments in the time horizon defined by the research design. Without the proper mechanisms of assessing causal inference, it can be easy to confound events with political fluidity and importance with cathartic moments that change dramatically the pace and direction of policy.

An *event* is defined as a spatially and temporally bounded happening that can be compared across cases and it is different from an occurrence that is a singular happening, specific to one case study at one point in time (Falleti and Mahoney 2015: 213). For CHA, one event necessarily shares the same characteristics as others and can be compared across different contexts to assess different reactions. In sum, an event is a significant but not determinant causal or data process observation in the language of social research methods. An event can become a critical juncture only if it produces a chain reaction effect with increasing returns over time that consolidate the outcome for a long time. There are events that become important within a sequence of changes, but do not necessarily trigger path-dependent trajectories. In fact, a long line of CPO events could represent just a different type of change, probably of the slow-moving sort (Mahoney and Thelen 2010).

A *sequence* is the temporal order in which events happen over time. For CHA sequencing becomes important when the order of earlier events is causally connected to the outcome of interest (Falleti and Mahoney 2015: 217). Even if sequences do not determine causation, there is an important argument to be made about sequences explaining long-term cycles of policy

that follow domestic or international economic cycles. For one thing, the order in which events happened greatly influences further developments. For another, the sequence also locates policies and policy change within the bigger context of case studies that have parallel sequences of events over macro issues that purportedly have direct impact over the policy issue the researcher is investigating. For example, macroeconomic policy changes from one paradigm to the next can greatly influence the sequence of changes of social policies, especially basic services like health and education. In this sense, the unfolding of correlated sequences can become CPOs within a different policy analysis that involves different policy issues.

Sequencing CPOs is one of the best tools to organize CPOs after the phase of process tracing has concluded. Assessing events and observations according to the order in which they appeared helps the researcher to identify missing data points, relevant information for segments along the sequence, or the need to improve their knowledge of certain events with in-depth interviews, variables and correlated information to ascertain causality with higher levels of confidence.

However, defining time frames with precision is not an easy task. Initial archival research and the revision of the existing literature can provide a rough time indication where the point of departure should be during the initial phases of research design. After the phase of primary and secondary data collection has concluded, time horizons in comparative policy analysis need confirmation and corroboration with other qualitative methodologies like in-depth interviews with key decision-makers along the sequenced timeline. More importantly, the time frame of analysis needs to connect perfectly with the results of process tracing: causal process observations.

There are many questions at this point of the research inquiry. How to evaluate all the elements so far, events, sequence, variables and assess the validity of the process tracing exercise? Is the time frame coinciding with the policy cycle? Are critical junctures correctly identified? Is it attributing correctly a paradigm shift? And, ultimately, is the process path dependent, linear or slow moving? This chapter contends that an index of policy intervention can be a useful tool to clarify all these questions.

4. INDEXES AS TOOLS OF ORGANIZING AND VALIDATING RESULTS

The index of policy intervention is a methodology to disaggregate policy components and track their evolution at regular intervals during the period under study. The index can work individually with quantitative or qualitative variables or can combine both. The only difference is that quantitative data entries will reflect indicators usually found in policy instruments like amounts of investment in education, health, or credit; while qualitative data entries replace facts or CPOs with points that vary in accordance with changes in the policy observation, like scorecards. Examples of qualitative information are laws and the different reforms made to them along the timeline; the existence of state agencies in charge or not, their changes over time recorded as scorecard numbers to register variations.

Indexes have been used in economics and social policy to measure impact (Angulo, Díaz, and Pardo 2016; Betanzo-Quezada, Romero-Navarrete, and Obregon-Biosca 2013; Lora 2001). And more frequently, management studies use them to generate scorecards of performance and company indicators (López, García, and Rodríguez 2007). Nevertheless, these

indexes are built as formal models of regression analysis with the ultimate goal of assessing the impact of different variables using only quantitative methods. The indexes are the only output of the study. In qualitative methods, the goal is to find the correct combination of variables and causal mechanisms. Theory-building is the final output. Data entries are not the only elements of analysis; there are also non-quantifiable kinds of information from agents and institutions that become causal process observations. Quantitative data is not omitted, but becomes part of the story. In qualitative approaches like CHA, an index is an innovation to increase accuracy in theory-building and theory-testing when analyzing long periods of time.

An index accomplishes three objectives: it illustrates trends in policy implementation and execution; it assists in determining types of policy change (slow motion, path dependency, regular intervals, cycles); and, fundamentally, it validates the time horizon of research and helps in identifying possible gaps or correcting misidentification of policy cycles.

To capture the different dimensions of policy change, the index depicts variations in policy configurations across case studies over the time horizon of the analysis. To work, the same data and institutional information should be available for the cases under study. This longitudinal analysis has the advantage of providing not only a trend line to understand the overall direction of changes, but also the scope or strength of policy intervention at different points in time, or when an event becomes a critical juncture.

The longitudinal analysis provides an overall view of the patterns of change, that is, the extent to which policies or segments of them were maintained or reformed over time. Then, using the components of the index, it becomes easy to identify policy configurations like path-dependent trajectories, periods of slow change, irregular intervals, and complete or incomplete cycles of policy.

Finally, the index also helps to increase internal validity of the time frame, since it could test within the same case studies but in different time frames, using the same variables and measurements applied to the first time frame.

4.1 How to Build an Index

The first step to build an index is to identify policy variables that are going to be process traced along the time frame defined in the research design. Then, quantitative and qualitative type of variables should be transformed into scores, as proxies of policy interventions along the timeline. For parsimonious results, the score registered for every policy component (CPOs) at regular intervals or data points along the defined time frame guarantees that even slight policy iterations will be regularly noted.

In a regular index that goes from 1 to 10, or from 1 to 100, the values assigned to each policy component should reflect the weight of the component within the overall policy design. As shown in the case below, tariff protection and credit were paramount indicators of infant industry protection policies. Thus, the compound value for both represents 40% of the final score. In some other cases, policy components can have equal weight explaining policy outcomes and trajectories of change. The different mixes of policy should be reflected in the weight every component is assigned with respect to the final score.

However, the most important aspect to consider is to clearly establish how points are assigned in each category. It is recommendable to establish scores in a continuum of possible indicators or CPO milestones, so as not to leave gaps or inexplicable leaps from one scenario to the next. The idea of a continuum is important to avoid random assignment of points that

can have problematic consequences in assessing the validity of the exercise. The idea is that each policy component has a rank of options or – in the case of numeric values – a rank of increasing or decreasing possibilities. The score points should reflect the ideal indicators of the paradigm that is subjected to a process of change.

The case below exemplifies this idea.¹ The maximum points are assigned to import substitution industrialization (ISI) variables because that is the paradigm that is in the process of being replaced. Thus, each policy component should have a ranking of points assigned according to the initial policy design under investigation.

5. CASE STUDY: ASSESSING INDUSTRIAL POLICY USING AN INDEX OF INTERVENTION

There were significant changes in the Latin American political economy after 2002, when a sustained boom of commodity prices in the region's main staple exports provoked a reconfiguration of political regimes that put to test the neoliberal reform agenda of the 1990s in South America. It was assumed that structural reform and a restructured role of the state in the economy was all but sanctioned. Yet, protectionist policy agendas in Brazil, Argentina, Venezuela, Bolivia, and Ecuador made a comeback after the year 2000. Among them, Ecuador took the most dramatic turn in 2008, announcing a full implementation of ISI policies. This puzzling reversal of reform invited a research inquiry into whether deep neoliberal reforms at the sectoral level even took place.

Besides macroeconomic restructuring, in compliance in most Latin American countries, there were some signs that sectoral and institutional reforms were lagging behind (Grindle 1991). Eduardo Lora (2001, 2007) advanced a scorecard of the state of reform across various issues and assessed each in a rank from 0 (no reform) to 1 (full reform achieved). Most countries fell between 0.6 and 0.7. There was no reason to believe industrial policy would fare better. After all, industrial policy was not just any policy. Latin American countries had embraced ISI since the beginning of the twentieth century, as the paradigm of choice to achieve economic development (FitzGerald 1998; Hirschman 1968; Prebisch 1949).

Central questions to address this puzzle were: What happened with industrial policy under neoliberalism? How and under what conditions did the transition from one paradigm to the next take place in late industrializers? What explains the sudden return of ISI to one of the members of the “pink wave”?

CHA was the best approach to answer these questions. Industrial policy and the developmental state literature have a long tradition using CHA as their preferred research method. Accounts of the political economy of Asian development by Amsden, Wade, and Johnson became seminal examples of how interconnected political, economic, and social factors explain the industrialization success in that region (Haggard 2015). They demonstrated how limited single causal explanations used by neoclassical economics or by rational choice political scientists could not explain long-term policy frameworks that went beyond drastic changes in rates of capital investment. By the same token, the same literature had not used the same elements to explain failed cases of industrialization, only successful ones (Ocampo 2002). Among failed cases of industrialization, late adopters were mostly left aside, especially in Latin America.

This logic of inquiry also demanded a small-N case study to trace the different elements and variables in the process of change. The Ecuadorian case of almost full reversal to ISI policies after 2008 provided an incredible opportunity to test the method of difference where similar cases that presented different outcomes in the dependent variable are paired for controlled comparison purposes.

Within late industrializers, Peru constituted an ideal match. For one thing, both Ecuador and Peru experienced the peak of ISI around the same time in the early 1970s, so they share the point of departure for the process of unraveling of ISI. On the other hand, both countries had similar political experiences with authoritarianism. These are the only two countries in the region where left-leaning nationalist military rulers embraced ISI as the core of their developmental objectives. Democratic governments carried out neoliberal reforms in both countries even though their democratic transition became extremely volatile with fragmented party systems. Additionally, both countries were subject to the same external pressures and changes in the international economy such as the boom in oil prices in the 1970s and then the sudden burst of commodity exports during the early 1980s, that affected domestic incentives channeled to manufacturing development. Most importantly, what makes these cases even more compelling for comparison is the variation of the dependent variable, in other words, how divergent became the paths they took after 30 years of neoliberal influence.

Process tracing was the method of choice to track back the sequence of causal mechanisms behind the transition from one paradigm to the next. Once all information was gathered and CPOs established, this inquiry advanced an index of intervention as a methodology to organize all the findings and make sense of the trajectory of 40 years of policy transition.

5.1 Components and Values of the Index

The policy components of the index are widely accepted central features of successful industrialization processes in East Asia and Western Europe, that are usually not part of a general macroeconomic program (Amsden 2001; Evans 1995; Wade 2004). These include instruments such as credit, and a tariff protection scheme exclusively dedicated to boost manufacturing; institutions like a central steering agency in addition to other dedicated institutions to help manufacturing development; and, importantly in Latin America, a legal framework that enforces the protectionist architecture. The caveat for this selection of variables is that they are minimum requirements of infant industry protection. Other qualitative elements of policy, like active discretionary power in the hands of specialized agencies and leadership or public-private partnerships, were not included (Wade 2004).

The overall score is built upon assigning values to variations in the three main components of industrial policy: legal framework (a maximum of 3), institutions (a maximum of 3), and instruments disaggregated in credit and tariffs (a maximum of 4). Table 7.1 summarizes the numbers assigned to each policy variation for each policy component that could add up a final score of 10. Each component is assessed along regular policy intervals of five years in the period between 1980 and 2010 to assemble the index. Thus, the index tracks changes since 1970 at regular intervals of five years each until 2010, keeping in mind that neoliberal reforms are expected to kick in from the interval of 1980–1985, which is also the first period of democratic rule in both countries.

For this case, the index goes from 0 to 10, where 10 could be the maximum aggregate score. A final score of 10 is only assigned to a complete protectionist industrial policy framework,

including ISI measures and infant industry protection for strategic or advanced-technology industries. At the other end of the spectrum, a final score of 0 represents no industrial policy interventions of any type. However, it would be difficult to report countries with no policy interventions at all. Even under the Washington Consensus guidance, industrial manufacturing has to benefit from the so-called “Productive Development Policies” (PDPs) (Melo and Rodríguez-Clare 2007) or horizontal industrial policies that promote “systemic competitiveness” (Hausmann, Rodrik, and Sabel 2008). The variables included in the index are explained below.

The *legal framework* can receive a maximum of 3 points when the law promotes national manufacturing development and includes a comprehensive set of incentives and institutional steering bodies. This component gets 2 points if the law exists, but it is no longer enforced, or other laws stripped out important components, such as credit, tax, or tariff incentives. One point is assigned if there is a law in place but it is vague in nature, or it is aimed at production in general where manufacturing is treated like any other sector. In this sense, Latin America experienced a very different situation than East Asian economies. In this region, the law usually precedes any policy intervention, and it usually embodies the spirit and specifics of the policy itself. Legislation has historically provided a framework of protection for industrialists and the productive sector in general. Therefore, these intermediate policies would score 2 points or less by their liberalizing disposition.

Institutions became one of the main beacons of the era of ISI policies. At its height, ISI envisioned a set of state institutions to guarantee industrial transformation, from planning bodies and ministries to technical assistance agencies and public financial corporations, specifically to support manufacturing development. This component gets 3 points when this set of institutions exists and they are coordinated by a single steering agency, usually the Ministry of Industries or the Planning body. Two points are assigned if some institutions like a Planning board, a Ministry of Industries or a specialized credit institution exist, but there is no coordination. Only one point is given if there is minimal bureaucratic support, expressed in just one public agency in charge of industrial policy as was the case of Peru during 1940, following the creation of the Industrial Bank.

Instruments provided the last four points of our index, broken down into tariffs and credit. Traditionally, ISI policies cemented the use of tariffs and subsidized credit as two of their most important incentives to boost manufacturing production (Chang 1994; Wade 2004). A *tariff* protection scheme, sometimes including non-tariff measures as well, and designed to boost manufacturing can add 2 points to the index. If tariff protection for manufacturing exists, but it is casual, or just some products are protected, this component gets just 1 point. If trade liberalization had occurred, often with a flat tariff policy in place, a score of 0 will be attributed.

Regarding *credit*, if there is a guaranteed public fund specifically channeling credit to manufacturing under subsidized rates, this component gets a score of 2. If there is credit availability for manufacturing, but it is not subsidized, or it is scarce, it gets a score of 1. Scores below 1 mean that a thorough liberalization has occurred, and credit is available only through private market mechanisms, and therefore they are not meant as policy incentives. Table 7.1 presents a synthesis of the points attributed in the index.

It is important to stress that this index does not measure policy impact of effectiveness, just the existence or not of an industrial policy framework at different intervals during a 30-year period and its persistence over time. Assessing policy effectiveness involves a broader set of

Table 7.1 *Index metadata*

Component	Variables	Points
Legal framework (Max of 3)	A comprehensive law of industrial promotion is in place and enforced.	3
	A law of industrial promotion exists but benefits have been stripped down, or the existing law is not enforced.	2
	Law exists, but manufacturing is treated as any other productive sector. No specific incentives exist.	1
Instruments:	A tariff and non-tariff protection scheme are in place specifically to protect infant industry.	2
A. Tariffs	Some tariff protection exists for targeted manufacturing products.	1
(Max of 2)	Flat tariff liberalization is in place.	0.5–0
Instruments:	Public funds specifically channeled to manufacturing exist; credit is subsidized.	2
B. Credit	Public credit exists but it is not subsidized.	1
(Max of 2)	Credit is available through private market mechanisms.	0.5–0
Institutions	The state has in place a specific set of institutions to promote manufacturing under a clear steering agency.	3
(Max of 3)	Institutions to support manufacturing exist, but there is no clear steering agency.	2
	At least one state agency supports manufacturing.	1

variables, including the impact of external shocks as well as the interaction of macroeconomic policy with targeted sectoral policy instruments.

5.2 Applying Values to Index Concepts and Indicators²

The index exercise provided an important picture of what happened with industrial policy from its peak during the 1970s and the 40 years of phasing out of its principal incentives. Tables 7.2 and 7.3 show the scoring for each macro-component as well as the total score. At the beginning of the period under study, military administrations in both countries scaled up a strong set of industrial policies. In both cases, a protectionist industrial law encompassing the different aspects of policy instruments and institutions providing support for manufacturing development was in place and fully enforced. The state created institutions to support it and they were fully funded, albeit with little or scarce coordination among them, a tariff framework was in place and a public fund – established in each law of industrial promotion – financed manufacturing ventures through subsidized credit. The similarities produce a departing score of 9 for both countries.

However, the momentous convergence ends there. As Tables 7.2 and 7.3 show, the pattern of differences starts quite early with the legal dismantling of the Law 18350 that boosted ISI in Peru and a sharp reduction in tariff protection and subsidized credit. Ecuador's changes lag behind in all components, except for a mild tariff reform during the 1980s.

The index of policy intervention confirmed the evidence coming from primary and secondary sources and interviews conducted in both countries (N=58). Indeed, Peru and Ecuador followed very different patterns of institutional change and the divergent trajectories can be seen in Figure 7.1.

Using the index, the path-dependence effect in the case of Peru becomes self-evident. On the other hand, a cyclical pattern – first of continuity of protectionist policies and then of change and reform – is also clear in the Ecuadorian case. The temporal sequence started in Peru with Velasco Alvarado's forceful intervention in 1968, and the sweeping institutional changes that were implemented added to the character of his impositions, creating a type of critical

Table 7.2 Ecuador's index of intervention

Years	Legal framework	Institutions	Tariffs	Credit	Total
1980–1984	3	2	2	2	9
1985–1989	3	2	1	1	7
1990–1994	2	2	0.5	0.5	5
1995–1999	2	2	0.5	0.5	5
2000–2004	2	2	0.5	0.5	5
2005–2009	3	2	1	1	7
2010–2015	3	2	2	2	9

Table 7.3 Peru's index of intervention

Years	Legal framework	Institutions	Tariffs	Credit	Total
1980–1984	3	3	1	2	9
1985–1989	1	2	1	1	5
1990–1994	1	1	0.5	0.5	3
1995–1999	1	1	0	0	2
2000–2004	1	1	0	0	2
2005–2009	1	1	0	0	2
2010–2015	1	1	0	0	2

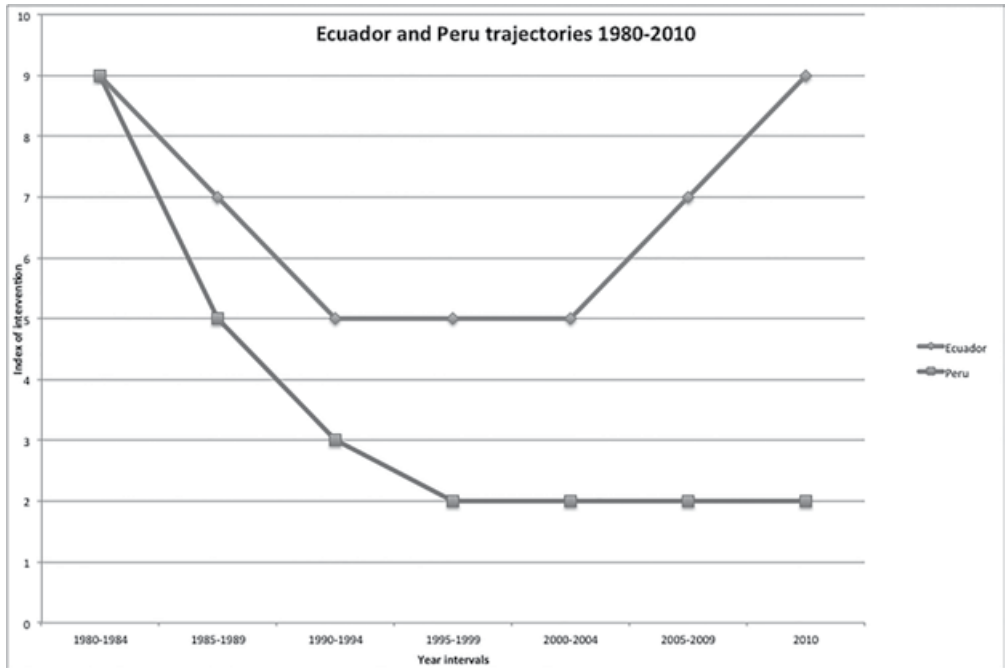


Figure 7.1 Differing patterns for Ecuador and Peru

juncture of dramatic remembrance for industrial policy. The neoliberal changes that followed his administration and that of his successor determined an institutional pattern of liberalization that became locked-in and self-reinforced over time. In other words, Velasco Alvarado's brand of industrial policy set the stage for a sustained and self-reinforced pattern of changes of the reactionary type (Mahoney 2000: 507). The reaction to Velasco's experiment grew stronger over time. Each new reform in the sequence leading to the dismantling of industrial policies is in part a direct reaction to Velasco's original intervention.

Contrarily, Ecuador did not experience such a sudden and radical moment. Of course, Guillermo Rodríguez Lara and the military junta that succeeded him became a critical juncture in the ISI policy history of Ecuador. But their actual accomplishment was due not to a change in policy course but to a windfall of oil revenues after discovering large reserves in the Amazon in 1973. The windfall just made the ISI package stronger and the cumulative peak in a long trajectory of protectionist layering incentives to develop manufacturing. This finding is consistent with Pierson's (2004) observation that consistent changes in one direction – in this case towards protectionist incentives – reached over time a cumulative threshold effect that usually signals a tipping point where change is expected (Pierson 2004: 83). Thus in Ecuador, the unwinding of reforms took the form of receding waves that slowly transformed the institutional framework. But they were not strong enough to prevent a backlash to return to the long-standing industrial protection schemes the state provided.

The 1980s marked a dramatic divergence in policy paths for Ecuador's and Peru's industrial policy. Peru started its departure from ISI protectionist policies as quickly as the democratic era started, reining in the policy instruments that were subsidizing national manufacturing through tariff protection and moving decisively away from state-directed manufacturing development towards a more market-oriented approach. By contrast, Ecuador did not unwind the ISI model for a decade, despite numerous economic crises that put its fiscal limits to the test. Even when both countries flirted with heterodoxy in the case of Peru from 1985 to 1988 and with orthodoxy in the Ecuadorian case from 1984 to 1987, a long-term view of the trend in both countries confirmed the divergency of policy initiatives for manufacturing, regarding instruments and institutional frameworks.

The departure deepens during the 1990s and 2000s, when Peru consolidated a neoliberal model of export-oriented development policies while Ecuador had an uneven and conflictive experience with the same reforms, only to end up reviving the ISI model when a left-leaning coalition won the presidential elections in 2006.

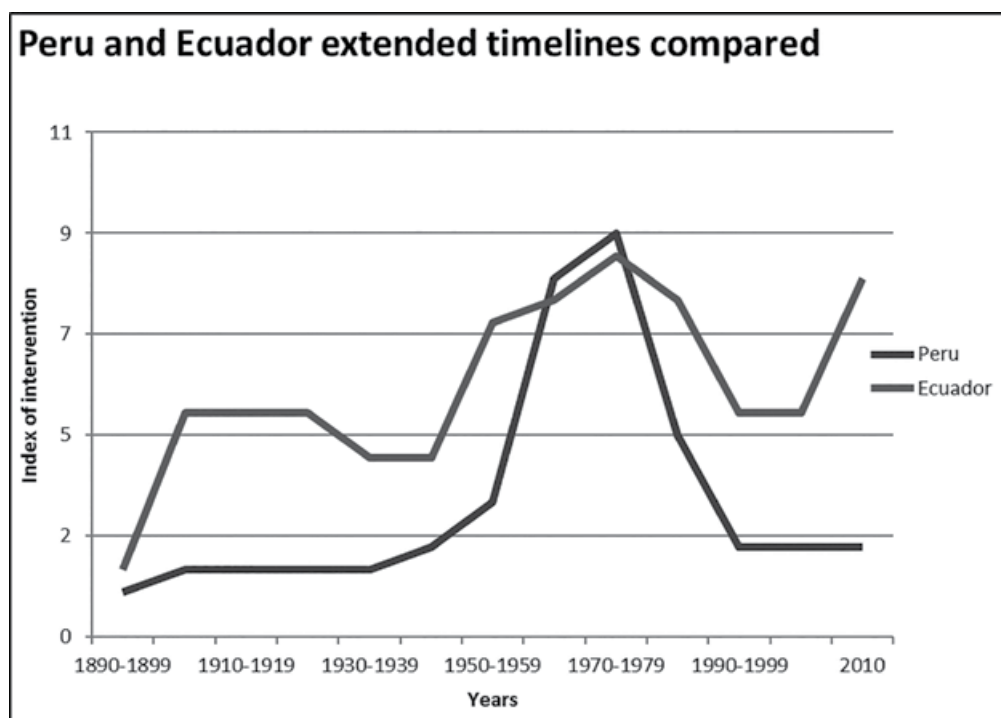
The trends highlighted by the index 1980–2010 raised the possibility that longer patterns of institutional evolution could be at stake in these two cases. The probability that institutional legacies could also play a role in the differential patterns of policy change under neoliberalism became another puzzle in the reconstruction of historical trends.

5.3 Checking Internal Validity

Cases like Peru's path-dependent trajectory are rare in the literature. Positive feedback is frequent, reaction is not. The inference about path dependency of the reactionary type needed to be double checked for internal validity. After all, the accounts of neoliberal reform always took President Fujimori's reforms in the 1990s as the starting point of positive feedback for the neoliberal dismantling of ISI. Notwithstanding, process tracing analysis evidenced otherwise

with respect to industrial policy. Among industrialists, politicians, technocrats, and even beneficiaries of ISI, reaction was well under way before Fujimori took power.

The index provided the possibility to test results extending the time horizon to engage more deeply in within-case analysis. Why not test the index since the beginning of modern industrialization that, according to most scholars, started in the last decade of the nineteenth century in Peru (Chiriboga 2013; Luna 2013; Pease-García and Romero-Sommer 1986, 2013; Thorp and Bertram 1978; Wils 1979)? Using the same variables and scoring categories explained above, the index was applied to the extended time frame with the only difference of longer time intervals of 10 years instead of 5 to facilitate visualization of long policy patterns. Even with the caveat that only secondary sources and available statistics informed the scoring exercise, not in-depth interviews, the extended index served its purpose. The long-term trajectory of industrial policy in both countries became clear and it confirmed the type of cycles and patterns that were only apparent with the original time frame (Tables 7.4 and 7.5).



Note: Initial time frame: 1980–2010.

Figure 7.2 Peru and Ecuador indexes compared (historical view 1890–2010)

The illustration of the extended index (Figure 7.2) provides a clear visualization. The most important finding out of the extended index is the fact that patterns of change after the 1980s closely resemble historical patterns and institutional traits present in these two countries since modern industrialization began. Peru’s industrial policy incentives have been historically minimal, scattered and without the force of an ISI package. The military government of

Table 7.4 Ecuador's expanded time frame index (using ten year intervals)

Years	Legal framework	Institutions	Tariffs	Credit	Total
1890–1899	0	0	1.5	0	1.5
1900–1909	3	0	2	0	5
1910–1919	3	0	2	0	5
1920–1929	3	0	2	0	5
1930–1939	2	0	2	0	4
1940–1949	2	1	1	0	4
1950–1959	3	1.5	2	0	7
1960–1969	3	2	2	0.5	7.5
1970–1979	3	2	2	1.5	8.5
1980–1989	3	2	1.5	1	7.5
1990–1999	3	1	0.5	0.5	5
2000–2010	3	2	2	1	8

Table 7.5 Peru's expanded time frame index (using ten year intervals)

Years	Legal framework	Institutions	Tariffs	Credit	Total
1890–1899	0	0	1	0	1
1900–1909	0	0	1.5	0	1.5
1910–1919	0	0	1.5	0	1.5
1920–1929	0	0	1.5	0	1.5
1930–1939	0	1	0.5	0	1.5
1940–1949	0	1	0.5	0.5	2
1950–1959	1	1	0.5	0.5	3
1960–1969	3	1.5	2	1.5	8
1970–1979	3	2	2	2	9
1980–1989	1	2	1	0.5	4.5
1990–1999	1	1	0	0	2
2000–2010	1	1	0	0	2

Velasco Alvarado dramatically changed the trajectory pushing a coercive ISI set of policies that included state-led industrialization, widespread discretionary nationalizations and expropriations and communal property of industries over time. In less than a decade Alvarado's policy backfired. As a result, the strong movement towards ISI and state-driven manufacturing development would likely become a cataclysmic event to the main beneficiaries of this set of policy interventions. They reacted forcefully: first dismantling the overall set of trade protections and then, the industrial protection law itself, once democracy was reinstated. The long trajectory showed why this was the case. The strong counter-reaction arrived soon and remained strong for the decades to come.

5.4 Explaining Causal Mechanisms

Three interconnected causal mechanisms explain the divergence of trajectories: institutional legacies, coordination among actors, and economic distribution of power. Peru's long tradition of a minimal state contrasts with the embedded character of Ecuador's long tradition of legal protectionism, dating back to the Liberal Revolution. Peru's close policy coordination among stakeholders – state technocrats and business elites – differs from Ecuador's “winner-take-all” approach for policy making. Peru's economic dynamism concentrated in Lima sharply departs

from Ecuador's fragmented regionalism with competitive and non-cooperative political leaderships in the coastal and the highlands regions.

The index contribution to the study of industrial policy in Ecuador and Peru is striking. First, it demonstrates that time and sequence played a significant role in determining path dependence and the reactionary sequence that followed Velasco's critical juncture. Velasco's ISI experiment was unique and it started too late compared to its regional counterparts. Ecuador turned out to be a very early supporter of infant industry protection policies and even ISI. The strong protection laws were there from the beginning of modern industrialization efforts and incentives were never dismantled; always padded up over time. Small attempts at neoliberal reform during the late 1980s and 1990s promptly failed within a community of stakeholders historically committed to protection.

Second, agency was vital either for overcoming protectionism and the institutions that upheld it or to support them and prevent their dismantling. In Peru, actors mobilized quickly to enhance their veto power over the policy process, seizing every opportunity to eliminate previous rules and create new rules of their own making during the early 1980s. Over time, a consensus developed among industrialists and private-sector actors that gave them the upper hand in the policy-making process. The positive gains of an early legal reform in 1982 and the shocking experience of demand-driven incentives during the first Alan García administration reinforced their capacity to shape industrial policy. At the end of President Fujimori's decade in power, the economic elite had become a powerful veto player that prevented any reversal of reforms. In Ecuador, industrialists and long-term state bureaucrats worked to keep incentives in place even during periods of crisis. Even when institutions changed in name or form, industrial protection ran deep in the state's approach to reform during the neoliberal years. This also explains why Rafael Correa's revisiting of ISI during the commodity boom of the 2000s was so easy to embrace for all the actors involved.

Third, historical legacies played a significant role. Looking for answers to the divergent outcomes of Ecuador and Peru after their respective ISI experiences, the research shows the utility of building explanations using longer time frames. The long-term trajectory of industrial policy in Peru was a good predictor of what was about to happen after the 1990s reform and its consolidation during the 2000s. In essence, Peru had returned to liberal economic management, embedded since the beginnings of the modern Peruvian state at the turn of the twentieth century. Velasco was just a significant anomaly in Peru's extremely liberal state building. Direct business-sector intervention in economic policy making was already institutionalized before Velasco. Private-sector participation in economic or productive policy making was expected and encouraged up until Velasco seized power in 1968. These two factors clearly generated self-reinforcing dynamics and eventually a backlash to Velasco's policies, since the latter constituted a radical departure from what the actors had come to expect.

Long-term traits in the evolution of industrial policy became embedded in each case. In the case of Peru, there was not a tradition of institutional or legal protection for industrial manufacturing until the 1950s. As a result, the strong movement towards ISI and state-driven manufacturing development would likely become a seismic event even for the main beneficiaries of this set of policy interventions. A graphic visualization of this long-term trait is shown in Figure 7.3.

The Ecuadorian case produced more unexpected findings not anticipated during research design. Secondary sources about Ecuadorian industrial policy are not as abundant as those for Peru. Only the index, once finished, showed clearly the uneven character of Ecuadorian tran-

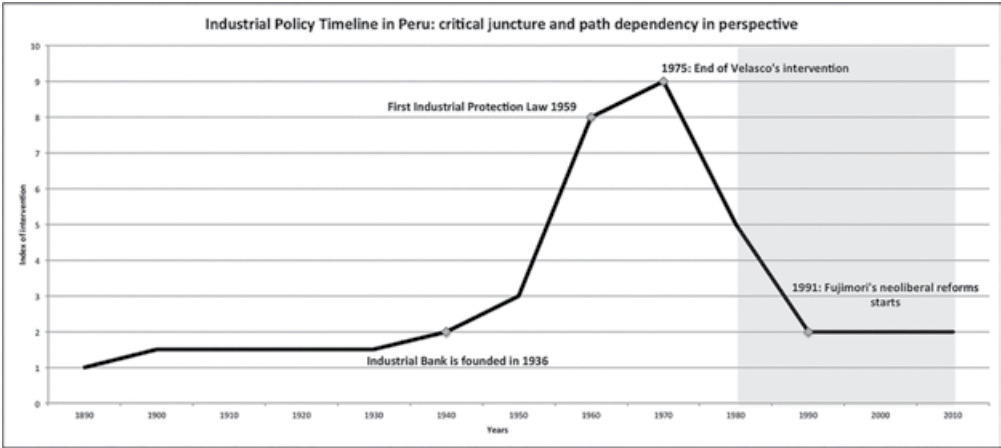


Figure 7.3 *Peru's neoliberal reforms in perspective 1890–2010*

sition to neoliberal productive policies and its condition as a late industrializer. Furthermore, the existing literature reached somewhat contradictory assessments about the timeline and features of the industrial policy framework (Chiriboga 2013; Oleas Montalvo 2013; Velasco 1981). For some scholars, industrial policy was non-existent until the windfall of oil revenue and the nationalistic appropriation of its revenues during the 1970s (De la Torre 1986; Oleas Montalvo 2013). Others argued that industrial policy did exist well before the 1970s, but the measures taken were insufficient to produce change and served the interests of oligarchical groups (Fischer 1983; Torres 2012). Thus, the most important finding was the realization that the legal framework for infant industry protection was set in place as early as 1906 and maintained ever since with different names or register numbers. With caveats, an industrial law protecting national manufacturing production was always at industrialists' disposal and scantily reformed even during the most difficult years of neoliberal reform like 1992–1996. Even some economic incentives stipulated by the laws like credit or tariffs stayed in place, albeit with upward or downward generosity depending on the booms or busts of international commodity prices, especially oil. In sum, the long-term institutional legacy was a sufficient but not necessary cause that explained Correa's revival of ISI.

As Figure 7.4 shows, the starting point of industrial protection in Ecuador was considerably earlier in the century and it kept moving upwards as time elapses. It is not surprising that when neoliberal reform arrived with force during the early 1990s, the political coalition that championed it advanced reforms only half way. By the end of the century, neoliberal reformism felt adrift in the midst of political turmoil. The second important element that explains this slow-moving, layering type of industrial policy change (Mahoney and Thelen 2010) is the entrenched divisiveness among regional economic elites which prevented any type of coalitional agreement regarding economic management in Ecuador, not to speak of shared vision of industrial development.

The long-standing distrust among economic elites translated into political action that enacted laws providing significant concessions and benefits to alleviate the ongoing strife. Looking from today's vantage point, it is not surprising that the Liberal Revolution, the first political attempt to create a unifying national vision (Ayala 1988), was also the one that passed

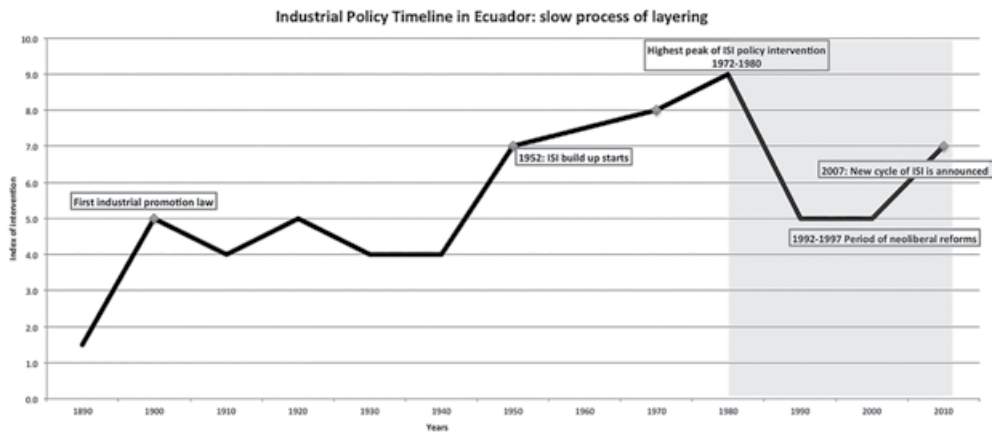


Figure 7.4 The long trajectory of industrial policies 1890–2010

the first piece of legislation to guarantee industrial protection and promotion as early as 1906. Industrial protectionist laws became a sort of guarantee for disgruntled regional groups in one cycle or another. Dismantling welfare gains from manufacturing factions at different points in time became as difficult as getting consensus for any other type of economic reform and only possible under hardship and deep economic crisis. Agency was simply not there to reform even when the existing legal framework became irrelevant under globalization and liberalization pressures during the last decade of the twentieth century.

Thus, agency was not a cohesive force in Ecuador, quite the contrary. On the one hand, business associations were always divided along regional lines. So much so that industrialists had formed chambers not along their sector's interest but along regional and even municipal lines. Coordination was never assured with the national government and usually respected bureaucrats and later technocrats had to become a third-party mediator among them. For that matter, Ecuador did not experience the type of critical juncture that Peru faced during Velasco Alvarado's tenure in power. General Rodríguez Lara never resorted to sweeping nationalizations, common to Velasco's brand of ISI policies in Peru. The harsh penalties, forced nationalizations and even the political exile of opponents were just absent in the Ecuadorian military approach to industrialists.

The index of intervention was key to understanding the nuance in the slow-moving changes that Ecuador experienced in its long-haul policy game. Disaggregating changes at the level of policy components, two elements became clear: ISI policies in Ecuador did change over time, but they did so in a gradual and cumulative way. Institutional *layering* (Thelen 2003) became the mode by which different actors negotiated their discrepancies without disrupting the existing institutional arrangements. Thus, new rules piled up on top of others. The period from 1980 to 2010 witnessed the creation of new institutions without dismantling previous ones.

6. CONCLUSIONS

This chapter demonstrates how important time frame definition is for policy analysis. Without a tool for rigorously validating time horizons of analysis when performing cross comparison policy cases, the potential risk of misrepresenting outcomes or critical junctures as determinants of the full policy cycle is always present. As the state becomes more and more complex, policy issues also become a moving target alongside the actors and institutions that are in charge of them. It is often the case that important issues like industrial policy or social compensation policies are administered by different agencies across time. Notwithstanding, change is slow and iterative. In other cases, the agencies do not change but policy shifts in paradigmatic ways and process tracing cannot help by itself to understand the different paradigms at play, absent an adequate consideration of comparative time frames in different case studies. As Falleti and Mahoney (2015) asserted, the sequence of events can be determinant to the pace that reform takes. The index of policy intervention has the possibility of aggregating the different aspects of policy in qualitative and quantitative indicators across long periods of time, so that dramatic changes or slow type evolution are more easily assessed.

Even when the index cannot and does not determine causality, it is a useful tool to organize, classify and understand process-tracing observations that in CHA are as numerous as any large-N study. It also helps to observe the strength of events to the extent that it codifies the extent of changes in incentives at any given moment in time, avoiding the tendency to attribute path-dependency trajectories to all critical junctures across policy cycles as the comparative case studies of Peru and Ecuador, detailed in this chapter, have shown.

The consequences for comparative policy analysis are clear. The index compounds quantitative and qualitative attributions that can be extremely helpful in the process of policy design and policy reform. The index provides a rapid visualization of long-term trends and can be easily linked to policy indicators to understand the correlation of both over time. Public administrators can also tailor the index to incorporate particular policy incentives either in quantitative terms (like bonuses, credit, tariffs, quotas, tax benefits, or cash transfers) or qualitatively measured in binary terms (the existence or not of: legal protection, agencies, consultative bodies, dedicated ombudsmen and the like). This small methodology has the advantage of bridging the gap between comparative policy scholars and public policy practitioners in understanding and improving public policy as a whole.

NOTES

1. The comparative case study presented here is a summary of the research scope and results of a doctoral dissertation covering 40 years of industrial policy reform in Ecuador and Peru. The original research detailed the amount, breadth and extension of policy incentives as well as decision-making processes through in-depth interviews of key actors during the time frame of analysis.
2. Laws, law-decrees passed during the military rule and other pieces of legislation came from the Justia database, retrieved at: <http://docs.peru.justia.com>. In the case of Ecuador, all laws, decrees, and supreme-decrees under the military rule were extracted from Registro Oficial, the official bulletin that historically has published all enacted legislation in Ecuador. Most of the cases require data collection from archival material rather than through the LEXIS database. All other information and data derives from Peru and Ecuador's respective Central Banks (Banco Central del Ecuador 1980–2010; Banco Central de Reserva del Perú 1934–2010, 2014). Additional data gaps were covered with World Bank statistics and indicators (World Bank 2014).

REFERENCES

- Amsden, A. (2001). *The Rise of “the Rest”: Challenges to the West from Late-Industrializing Economies*. New York: Oxford University Press.
- Angulo, R., Díaz, Y., and Pardo, R. (2016). The Colombian multidimensional poverty index: measuring poverty in a public policy context. *Social Indicators Research* 127 (1): 1–38.
- Banco Central del Ecuador (1980–2010). *Memoria anual*. Retrieved August 1, 2018 from <http://repositorio.bce.ec/>.
- Banco Central de Reserva del Perú (1934–2010). *Memorias*. Retrieved September 1, 2018 from <http://www.bcrp.gob.pe/publicaciones/memoria-anual.html>.
- Banco Central de Reserva del Perú (2014). *Cuadro y anales históricos*. Retrieved September 1, 2018 from <http://www.bcrp.gob.pe/estadisticas/cuadros-anales-historicos.html>.
- Barzelay, M. and Gallego, R. (2010). The comparative historical analysis of public management policy cycles in France, Italy, and Spain: symposium introduction. *Governance* 23 (2): 209–23.
- Bennett, A. (2010). Process tracing and causal inference. In H. E. Brady and D. Collier (Eds.), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield, 207–36.
- Betanzo-Quezada, E., Romero-Navarrete, J. A., and Obregon-Biosca, S. (2013). An index to evaluate public policies in urban freight transport. *Gestión y Política Pública* 22 (2): 313–54.
- Capoccia, G. (2015). Critical junctures and institutional change. In J. Mahoney and K. Thelen (Eds.), *Advances in Comparative-Historical Analysis*. New York: Cambridge University Press, 147–79.
- Capoccia, G. (2016). When do institutions “bite”? Historical institutionalism and the politics of institutional change. *Comparative Political Studies* 49 (8): 1095–127.
- Chang, H. J. (1994). *The Political Economy of Industrial Policy*. New York: St. Martin’s Press.
- Chiriboga, M. (2013). *Jornaleros, grandes propietarios y exportación cacaotera 1790–1925*. Biblioteca de Ciencias Sociales, 76. Quito, Ecuador: Universidad Andina Simón Bolívar Sede Ecuador-Corporación Editora Nacional.
- Collier, D. and Berins-Collier, R. (1991). *Shaping the Political Arena: Critical Junctures, the Labour Movement and Regime Dynamics in Latin America*. Notre Dame, IN: University of Notre Dame Press.
- Collier, D., Brady, H. E., and Seawright, J. (2010). Outdated views of qualitative methods: time to move on. *Political Analysis* 18 (4): 506–513.
- Conran, J. and Thelen, K. A. (2016). Institutional change. In O. Fioretos, T. G. Falletti, and A. Sheingate (Eds.), *The Oxford Handbook of Historical Institutionalism*. New York: Oxford University Press, 51–70.
- Evans, P. (1995). *Embedded Autonomy: States and Industrial Transformation*. Princeton, NJ: Princeton University Press.
- Evans, P. B., Rueschemeyer, D., and Skocpol, T. (Eds.) (1985). *Bringing the State Back In*. Cambridge: Cambridge University Press.
- Falletti, T. and Mahoney, J. (2015). The comparative sequential method. In J. Mahoney and K. Thelen (Eds.), *Advances in Comparative-Historical Analysis*. New York: Cambridge University Press, 211–39.
- FitzGerald, V. (1998). La CEPAL y la teoría de la industrialización. *Revista de la CEPAL*, Número Extraordinario. Retrieved September 1, 2018 from <http://repositorio.cepal.org/handle/11362/12123>.
- Gallego, R., Barbieri, N., and Gonzalez, S. (2017). Explaining cross-regional policy variation in public sector reform: institutions and change actors in the health sector in Spain. *Public Policy and Administration* 32 (1): 24–44.
- George, A. L. and McKeown, T. J. (1985). Case studies and theories of organizational decision making. *Advances in Information Processing in Organizations* 2 (1): 21–58.
- Grindle, M. (1991). The new political economy: positive economics and negative politics. In G. Meier (Ed.), *Politics and Policy Making in Developing Countries: Perspectives on the New Political Economy*. San Francisco, CA: ICS Press, 41–67.
- Haggard, S. (2015). The developmental state is dead: long live the developmental state! In J. Mahoney and K. Thelen (Eds.), *Advances in Comparative-Historical Analysis*. New York: Cambridge University Press, 39–66.
- Hall, P. A. (1986). *Governing the Economy: The Politics of State Intervention in Britain and France*. New York: Oxford University Press.

- Hall, P. A. and Taylor, R. C. (1996). Political science and the three new institutionalisms. *Political Studies* 44 (5): 936–57.
- Hausmann, R., Rodrik, D., and Sabel, C. (2008). Reconfiguring industrial policy: a framework with an application to South Africa. Center for International Development, Working Paper 168. Harvard University.
- Hirschman, A. O. (1968). The political economy of import-substituting industrialization in Latin America. *Quarterly Journal of Economics* 82 (1): 1–32.
- Levi, M. (1997). A model, a method, and a map: rational choice in comparative and historical analysis. In M. I. Lichbach and A. S. Zuckerman (Eds.), *Comparative Politics: Rationality, Culture, and Structure*. Cambridge: Cambridge University Press, 19–41.
- López, M. V., García, A., and Rodríguez, L. (2007). Sustainable development and corporate performance: a study based on the Dow Jones sustainability index. *Journal of Business Ethics* 75 (3): 285–300.
- Lora, E. (2001). Structural reforms in Latin America: what has been reformed and how to measure it. Working Paper #466. Washington, DC: Inter-American Development Bank Research Department.
- Lora, E. (Ed.) (2007). *The State of State Reform in Latin America*. Palo Alto, CA: Stanford University Press/World Bank/Inter-American Development Bank.
- Luna, M. (2013). *Orígenes de la política económica del desarrollo industrial en el Ecuador 1900–1960*. Quito, Ecuador: Ministerio Coordinador de la Política Económica.
- Mahoney, J. (2000). Path dependence in historical sociology. *Theory and Society* 29: 507–48.
- Mahoney, J. and Thelen, K. (2010). A theory of gradual institutional change. In J. Mahoney and K. Thelen (Eds.), *Explaining Institutional Change: Ambiguity, Agency and Power*. New York: Cambridge University Press, 1–37.
- Mahoney, J. and Thelen, K. (2015). Comparative-historical analysis in contemporary social science. In J. Mahoney and K. Thelen (Eds.), *Advances in Comparative-Historical Analysis*. New York: Cambridge University Press, 1–36.
- Mele, V. and Ongaro, E. (2014). Public sector reform in a context of political instability: Italy 1992–2007. *International Public Management Journal* 17 (1): 111–42.
- Melo, A. and Rodríguez-Clare, A. (2007). Productive development policies and supporting institutions in Latin America and the Caribbean. In E. Lora (Ed.), *The State of State Reform in Latin America*. Palo Alto, CA: Stanford University Press/World Bank/Inter-American Development Bank, 317–53.
- Moore, B. (1966). *Social Origins of Dictatorship and Democracy*. Boston, MA: Beacon Press.
- Ocampo, J. A. (2002). Trade policy and industrialization in Colombia, 1967–91. In G. K. Helleiner (Ed.), *Trade Policy and Industrialization in Turbulent Times*. London: Routledge, 150–187.
- Pease-García, H. and Romero-Sommer, G. (1986). *El ocaso del poder oligárquico: La lucha política en la escena oficial 1968–1975* (4th edition). Lima: DESCO.
- Pease-García, H. and Romero-Sommer, G. (2013). *La política en el Perú del siglo XX*. Lima: Fondo Editorial Pontificia Universidad Católica del Perú.
- Pedersen, S. (1995). *Family, Dependence, and the Origins of the Welfare State: Britain and France, 1914–1945*. Cambridge: Cambridge University Press.
- Pierson, P. (2004). *Politics in Time: History, Institutions, and Social Analysis*. Princeton, NJ: Princeton University Press.
- Prebisch, R. (1949). Growth, disequilibrium and disparities: interpretation of the process of economic development. *Economic Survey of Latin America, 1949-E/CN.12/164/Rev.1-1951*-pp. 3–85.
- Radaelli, C. M., Dente, B., and Dossi, S. (2012). Recasting institutionalism: institutional analysis and public policy. *European Political Science* 11 (4): 537–50.
- Skocpol, T. (1979). *States and Social Revolutions*. Cambridge: Cambridge University Press.
- Steinmo, S., Thelen, K., and Longstreth, F. (Eds.) (1992). *Structuring Politics: Historical Institutionalism in Comparative Analysis*. Cambridge: Cambridge University Press.
- Thelen, K. (2003). How institutions evolve: insights from comparative historical analysis. In J. Mahoney and D. Rueschemeyer (Eds.), *Comparative Historical Analysis in the Social Sciences*. Cambridge: Cambridge University Press, 208–40.
- Thorp, R. and Bertram, G. (1978). *Peru 1890–1977: Growth and Policy in an Open Economy*. New York: Columbia University Press.
- Wade, R. (2004). *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization* (2nd edition). Princeton, NJ: Princeton University Press.

- Wils, F. (1979). *Industrialization, Industrialists, and the Nation-State in Peru: A Comparative/Sociological Analysis*. Berkeley, CA: Institute of International Studies, University of California.
- World Bank (2014). *World Development Indicators*. Retrieved April 18, 2018 from <http://data.worldbank.org/data-catalog/world-development-indicators>.

8. Comparing international policy transfers

Osmany Porto de Oliveira

Yet it moves!

1. INTRODUCTION

How can we compare an object of public policy that travels across the world's national frontiers? The field of public policy transfer, diffusion, and international circulation has been experiencing dizzying growth lately, and has received attention in analyses of public policy and international relations.¹ The interest in this phenomenon has accompanied the contemporary empirical reality of globalization, in which ideas, institutions, and instruments of public action circulate more and more often between governments, accelerating this exchange of policy knowledge. This movement has been led by various actors, such as international organizations (Weyland 2007; Woods 2006), governments (Dolowitz and Marsh 2000), non-governmental actors such as think tanks (Stone 2001), epistemic communities and transnational networks (Dunlop 2009), private consulting firms and foundations (Porto de Oliveira and Pal 2018), and even individuals (Dezalay and Garth 2002; Porto de Oliveira 2017).

The phenomenon of policy transfer also presents distinct dynamics. Richard Rose (1991) has already insisted on the fact that political actors have intensified practices in which they nourish experiences from other countries. Such behavior in public policy making can produce copies, emulations, hybrids, inspirations, and mixtures of public policies coming from abroad. Many countries have engaged in policy transfers to other governments, through foreign policy and development cooperation practices, using this strategy as an instrument of soft power (Porto de Oliveira and Pimenta de Faria 2017). In addition to this, reactions against globalization have led to the emergence of localism, nationalism, and protectionism in distinct forms, and have drawn the attention of specialists to this. In this case, experts seek to understand why certain actors resist policy transfers, as can be seen in the recent case of Brexit (Porto de Oliveira and Pal 2018).

In spite of the growing number of studies of policy transfers, there has been little coordination and consistency in this field of study, which has so far produced very similar concepts and terms to describe the same phenomenon, which in simple words consists of the "traveling" of ideas, institutions, and political models.² In addition, the transversality of public policies and their growing profile in distinct areas of knowledge have contributed to a broader examination of this issue based on diverse disciplines such as: political science, anthropology, sociology, and geography to mention just a few. This emerging multi-disciplinary profile has on one hand enriched the debate, and on the other, has resulted in disputes in terms of how to understand this issue, which has led to research traditions within disciplines, as well as interdisciplinary disputes.³

In this chapter, following our proposal as outlined in another publication,⁴ to make it easier to understand this phenomenon, we will distinguish three types of movements in which

public policies “travel”: transfer, diffusion, and circulation. Policy transfer can be understood as a process through which policy objects move internationally in time and space and can produce some level of interdependence between actors. This means that the decisions previously taken by one actor affect others. To make this clearer, we can think of the social policies adopted in Brazil as Conditional Cash Transfers (CCT) which have influenced the implementation of similar instruments in Chile and in the Philippines (Howlett, Ramesh, and Saguin 2018; Osorio Gonnet 2019). The three terms mentioned before can be distinguished by the scope of these movements, with a transfer being a one-off transport of a policy from one government or organization to another. Meanwhile diffusion is wider and can occur in regional clusters, such as the adoption of state reforms in Latin America. Circulation, in turn, consists of a broader movement which involves more countries, model transformation processes, back and forth policy dynamics, and mutual learning relationships. The term transfer will be used in this text as a generic way to refer to this phenomenon.

International policy transfer occurs in the domestic and international spheres and is affected by the influence of actors who operate in both arenas. Observing the transnational dimension is almost inevitable and the analysis of these processes to a certain extent cannot be dissociated from the comparative task. In fact, policy transfer implies, in its process, looking at other experiences in terms of time and space. Policy transfer specialists focus on an object that does not have a well-defined territoriality, has transnational connections, moves across distinct jurisdictions, and can also be modified during the process by distinct involved actors according to their perspectives and interests. Despite the importance of this theme in contemporary public policy analysis, this field still does not have suitable methodological instruments to compare transfers. This is due above all to the fact that conventional methodologies are focused on comparing territorial units, and are hostage to “methodological nationalism” (Wimmer and Glick Schiller 2002).

Given this situation, we can ask a number of questions: How can we compare an object of study that does not have defined boundaries? What should be done when your object moves between cities and countries? What do we need to take into account to analyze an object that can transform itself over time and across spaces? How can we adapt conventional methods of comparative public policy analysis to study policy transfer? Which strategies can be used to understand its processes, causes, and effects? Finally, is it possible to speak of a methodological identity for the analysis of international policy transfers? This field adopts existing tools in the policy analysis area, adapting them and developing its own comparative methodological techniques and strategies to understand and explain international policy transfers. The literature uses the methodological dimension in a more or less intense and structured manner, which can vary according to each study. The argument of this chapter is that in order to access policy transfer in depth, we need to move beyond comparative public policy analysis centered on national frontiers. There exists a solid group of studies that makes it possible to identify the specificities of policy transfer comparisons. These studies exhibit great explanatory potential to examine this emerging policy phenomenon.

This chapter is divided into five parts, in addition to the introduction and conclusion. Section 2 deals with the centrality of the state in comparative studies and presents solutions that have been proposed in the policy transfer field, including a discussion of the Galton problem. Section 3 argues the need to move beyond international comparisons and include the transnational dimension in analyses. Section 4 presents the specificities related to the main issues, concepts, and objects of this field’s analysis. Section 5 deals specifically with the dis-

tinct forms of comparison which are possible to understand policy transfers. In section 6, we will synthesize all of this and present some challenges that still exist in this field in regard to comparative methodology, such as the issue of comparing transfer outcomes, and will indicate important directions for new studies.

2. THE COMPARATIVE ANALYSIS OF POLICY TRANSFERS: A CHALLENGE, SOLUTION, AND OPPORTUNITY

Analyzing diffusion processes presents a challenge, an opportunity, and a solution for the comparative public policy field. The challenge consists of the fact that conventional methodological strategies for comparing policies are focused on analyses that emphasize the study of delimited units. This places restrictions on objects whose transnational dimension is almost intrinsic. In fact, one of the most common ways of comparing a few countries is observing similar systems with different results (Most Similar Systems Design – MSSD) or different systems with similar results (Most Different Systems Design – MDSM). This method is based on proposals for correlations and differences presented by John Stuart Mill (1843) (Landman 2007: 70). The MSSD technique seeks to observe the characteristics that remain constant in a system to identify the factor which stands out in one or more cases. The identification of this distinctive factor makes it possible for the researcher to explain why these variations occur in the results. In turn, the MDSM model seeks to distill the explanatory factor in the internal characteristics of the countries.

Another way of comparing is to use case studies. According to Landman (2007: 28), case studies dealing with one country can be used to make comparisons when the purpose of the research is to confirm or refute hypotheses and theories which have been established in the literature, as well as to generalize and produce concepts that can be applied to a broader number of cases. In a distinct manner, John Gerring (2004) also argues that case studies can serve to make comparisons. In his opinion, a comparison occurs when you compare units within a case. However, the definition of a case has a limitation in terms of international policy transfer studies. This is due to the concept of the unit, which according to this author “connotes a *spatially bounded phenomenon*—e.g., a nation-state, revolution, political party, election, or person—observed at a single point in time or over some delimited period of time” (Gerring 2004: 342, our emphasis). The spatial restriction of the unit of observation is problematic in analyzing a transfer.

The internal/external dichotomy in the methodologies used to analyze public policies has been questioned in this field. In general terms, we can argue, in a stylized form, that the field of public policy analysis understands this phenomenon as occurring “within the state” and that the area of international relations is dedicated to understanding phenomena that occur “outside the state”. With the process of globalization and phenomena related to this process, internal/external boundaries have become more and more fluid, which makes it necessary to review conventional analysis strategies, which can also include the traditional dimension. Diane Stone (2008: 23) uses the aforementioned expression “methodological nationalism” to refer to this conventional approach of public policy analysis and sees it as an important barrier that the field needs to overcome. As we will see in detail later in this chapter, international policy transfer analysis may seek to understand the processes that lead one country or another to adopt a public policy model, the impact of the transfer on the policy’s success or failure, as

well as the circulation of the instrument itself from one country to another, and the actors that have intervened during this movement in transporting these objects in terms of time and space. In this sense, conventional comparative strategies by spatially delimiting the units of analysis, place an important barrier in the way of understanding phenomena like policy transfers, which involve elements, objects, or units that “have no boundaries” and that “travel” from one jurisdiction to another, transforming themselves along the way.

Policy transfer studies and the development of more robust methodologies can offer a solution to an inconvenient question for comparative specialists, known as “Galton’s problem”. This true “stone in the shoe” of comparative methodology consists of Galton’s finding at the end of the nineteenth century that correlations observed between social institutions are not just the fruit of the dynamic results of a society’s internal processes, but also may be the result of broader movements of cultural diffusion (Goldthorpe 1997). Galton’s ideas raise the question of interdependence between societies as a variable which can affect political processes and can have a direct impact on arguments related to autonomous institutional development. In the words of Peters, in terms of public policy, this problem is related to “the difficulty of identifying how much of the observed pattern of common administrative practices is a result of indigenous developments and how much is a product of transfer” (Peters 1997: 86).

Galton’s problem, therefore, insists on the difficulty or impossibility of isolating units, and imposes a limit on comparative social policy analyses, which specialists are always looking to overcome. This complication for the conventional methodology is part of the object of studying the diffusion of policies which are expressly designed to observe the dynamics of interdependence (Braun and Gilardi 2006) that exist when actors adopt ideas and institutions from other locations. To overcome the limitations of conventional strategies, the field of policy transfer has fortified itself with *sui generis* methodological techniques, often taking advantage of existing propositions in the public policy analysis area. Before covering the main comparative strategies, it is important to present how we will move from an international comparison to a comparison which can integrate the transnational dimensions of public policies.

3. BEYOND COMPARATIVE COUNTRY ANALYSIS: TRANSNATIONAL COMPARISON

Comparing the environment of the international transfer of ideas, models, and institutions faces a great challenge, which is the need to include transnational variables in the construction of public policies. This means solving Galton’s problem. It implies the need to adapt classical postulations in this field based on the frontiers of the state and the separation of units as a central element in political science comparisons. The purpose of this chapter is not to present a normative position in terms of comparative analysis, that is, imposing comparative public policy analysis strategies on the reader. On the contrary, it is a systemization of the field which observes what is being produced and presents, as if it were a compass, the forms of comparison used in research. This effort demonstrates the innovation of this idea. The principal element in this sense is the inclusion of the “transnationalization of public action”.

One way to handle this issue has been presented by Patrick Hassenteufel (2005). Hassenteufel shows the limits of international comparisons, arguing that “in making the issue of national specificities the central issue and emphasizing diachronic continuity, comparative works ... present a twofold limitation in neglecting convergence and minimizing change” in

public policies (Hassenteufel 2005: 121). Transnational comparisons are presented as a solution to the issue of convergence, or in other words, the adoption of similar public policies by a group of heterogeneous actors facing the same problems. The elements that should be included in the comparison to transport the analysis from the international to the transnational perspective are the following: (1) the role of transfer agents, their power relationships, the resources they have, and the spaces where they circulate; (2) the discursive strategies used to disseminate policies, such as the definition of exemplary cases and repetitive rhetoric about how to create these public policies and how not to, and which are presented above all by international organizations; and (3) the local reappropriation and attaching of local significance to models from other places, or in other words, translations that occur during the transfer process that make it possible to observe variations in practices, instruments, and institutions.

In an analogous manner, Jennifer Robinson, in speaking of urban policies criticizes conventional comparisons, pointing out that “the discreet territoriality of the conventional comparative imagination is a poor fit with complex, networked and dispersed spatialities of the interdependent urban territories ... which characterize contemporary urbanisation” (2018: 222). In her view, we need to “pay close attention to the array of transnational processes shaping distinctive policy outcomes and development paths as they come together in one specific place – to explore how ‘elsewhere’ is folded into localised growth paths” (Robinson 2018: 223). Robinson also points out informality and other correlated dynamics which are characteristic of transfer processes, remembering that frequently “ideas and practices arrive from elsewhere or emerge in particular contexts in all sorts of ways – through forgotten conversations at meetings, the long distant reading of publications or reports, unpredictable friendship and collegial networks, as well as formal or informal associations, in which taken for granted understandings might be confirmed” (2018: 233).

Paying attention to the transnational dimension has become a distinctive element in the comparison of policy transfers and, to some extent, an imperative to researchers who seek to explain this phenomenon. In general, studies within the transfer field of research are dedicated to a middle-N number of comparisons.⁵ There is also a significant amount of research dedicated to case studies. We will consider also as comparisons, case studies that involve more than one unit, in accordance with the proposals of Gerring (2004) and Landman (2007). In this chapter we will concentrate on studies that emphasize qualitative strategies. It is important to note that there is no purity in comparative analyses, given that comparing processes and intervening variables or results is a choice made by each specialist in an attempt to find the best response to one or more research questions. The studies examined here do not reflect the totality of the area, but make up an important group of articles selected by the author. In the next section we will introduce in a concise form the main research issues, questions, and concepts before beginning our systematization of the methodological strategies used by authors in this field.

4. ISSUES, QUESTIONS, AND CONCEPTS FOR POLICY TRANSFERS COMPARATIVE ANALYSIS

The object of analysis of policy transfer is the displacement of policy ideas, models, and institutions in time and space. Researchers in this area have dedicated themselves to the analysis of various empirical objects such as democracy (Huntington 1993), markets (Simmons, Dobbin, and Garrett 2008), state reforms (Weyland 2007), social policies (Osorio Gonnet

2019; Saburin and Grisa 2018), participatory institutions (Porto de Oliveira 2017), and judicial norms (Dezalay and Garth 2002), among others. Over time this field's literature has developed its own analytical frameworks, questions, concepts, and strategies which have progressively established its identity as an area of study.

Dolowitz and Marsh's analytical framework (2000) was one of the pioneering proposals and it is a thorough presentation of the elements that researchers should take into consideration to understand policy transfers. In their framework, this phenomenon can be studied as an independent variable, that is, as the transfer process, or as a dependent variable, which consists in the observation of the transfer's impact on the "success" or "failure" of the policy. Most of the field has concentrated its analyses on the transfer process dimensions and the variables that intervene during this process. Dolowitz and Marsh (2000: 8) have developed a group of questions to understand the phenomenon under examination: (1) Why do these actors engage in policy transfers? (2) Who are the main actors involved in transfer processes? (3) What is transferred? (4) Where are the lessons drawn from? (5) What are the different levels of the transfer? (6) What restricts or facilitates the transfer process? (7) How are transfer processes related to the "success" or "failure" of these policies? This analytical framework has been widely used in this field, but has also been criticized for its high degree of formalization and linearity. These critiques are based on the observation that transfers are not rational and mechanical, but rather involve superimpositions, multiple time frames, and the transformation of an object in movement. In terms of these criticisms, this model is interesting to initiate research and design methodologies, and it can be complemented or adjusted with more complex approaches to policy transfers which will be presented below.

The literature produced in France inspired by the legacy of Bruno Latour and the sociology of public action approach, insists on the role of translation, or in other words, the metamorphoses of objects in the appropriations of the actors and the reformulation of the policy's significance when it arrives in other places.⁶ The cognitive dimension of public policies can also be appreciated in the French debates which pay particular attention to the role of the underlying ideas, ideologies, and meanings of the models in circulation (Bruno, Jacquot, and Mandin 2006; Hassenteufel and de Maillard 2013). Recently Porto de Oliveira and Pal (2018) further broadened this array of questions by including the dimensions of the resistance to processes, the power relationships among actors, private actors, transnational arenas where agents publicize their models, and finally the directions of transfers, which originally were observed through the flow of policy from the North to the South, but which now also involve South–North, South–South, and East–West transfers, etc.

The most recent debate has included attempts to integrate process micro-dynamics in the macro-sociological analysis (Hadjiisky, Pal, and Walker 2017). The combination of the observation of world trends – such as the development of United Nations agendas, international regimes, and global governance processes – and their implementation through transfer processes on the local level is a way of combining both of these scales. In our previous study about the international diffusion of Brazilian participatory budgeting (PB), this trans-scalability was organized through the temporal reconstruction of the exiting of a public policy from an exclusively domestic environment and its entering the international agenda dealing as well with transfers to certain regions, countries, and sub-national units (Porto de Oliveira 2017). The micro-scale also can refer to more direct and specific observations made by individuals who participate in the processes and internal dynamics of international organizations and other institutions.

The causality dimension is also important to explaining the phenomenon of policy transfers. There is a consensus in the field that there are four causal mechanisms or forces that intervene in diffusion, facilitating or stimulating the policy transfer: coercion, learning, emulation, and competition (Graham, Shipan, and Volden 2013; Simmons, Dobbin, and Garrett 2008). These mechanisms can be expanded to other relationships that induce diffusion such as international cooperation, the circulation of individuals, networking, and other forces (Porto de Oliveira 2017). Recent studies have sought to identify the mechanisms that complicate, obstruct, and limit transfers such as technical complexity, political and organizational culture, and resistance to coercive designs, counter-hegemonic models (Hadjiisky, Pal and Walker 2017: 16), political changes in government (Porto de Oliveira 2017: 88), and budgetary restrictions. However, there still is no organized, dense body of work regarding the mechanisms that constrict policy transfers.

A comparison of policy transfer serves not only to generate good descriptions of its contexts and processes, but also to classify objects, processes, and phenomena, test hypotheses, and validate, refute and generate new theories, using or producing concepts and identifying regularities or causalities to explain the phenomenon in question. There is still much to be explored in the environment of the transnational dimension of public policies, and looking at transfers and producing descriptions is very important so that we can access this dimension and understand its *modus operandi*. In the next section we will present a group of important strategies which have been used by authors in this area to produce comparisons.

5. WHAT AND HOW SHOULD WE COMPARE IN INTERNATIONAL POLICY TRANSFERS ANALYSIS?

Studies of policy transfers can be dedicated to processes, results, independent variables, and also the transnational movement of the object transferred itself, or in other words the instrument of public action which has been adopted in distinct locations. Among the elements that can be considered in comparisons are: (1) variations or translations of models; (2) the transnational root (when we are dealing with an “import” process) or the transnational destination of the policy (when we are dealing with an “export” process); (3) the power relationships and role of transfer agents and their behaviors; (4) the mechanisms that facilitate or constrict the adoption of models; and (5) the *modus operandi* of the arenas. Insofar as this is an emerging field, there are still many elements that are not clear in terms of how to compare them. In this section we will use policy transfer studies to present distinct forms of comparison, with the first being the analysis of the process, and the second the analysis of independent elements. There are still few studies dedicated to comparing the influence of transfers on policy results, and this is why we will reflect briefly on this aspect in the conclusion as a challenge to be overcome. Authors of the studies described below do not always present an extensive and detailed methodology section, which is why the field systematization reproduced here uses studies as a source of inspiration to elaborate propositions, which can encourage readers to venture into the comparative strategy for international policy transfer analysis.

5.1 Comparing Transnational Processes

Reconstructing the “paths” of a transfer, observing the roles of actors, causal mechanisms, and macro-, medium, and micro-scale dynamics is one of the most common analytical practices. The technique of process tracing is often used by these authors. According to George and Bennett (2005: 6) process tracing “attempts to trace the links between possible causes and observed outcomes”. Using this technique, the researcher examines documents and primary and secondary sources “to see whether the causal process a theory hypothesizes or implies in a case is in fact evident in the sequence and values of the intervening variables in that case” (George and Bennett 2005: 6).

Through process tracing, authors have sought to understand the causal chains which lead to policy transfers. Studies aspire to identify tipping-points and processual changes through the observation of events (such as the Eco-1992, which took place in Rio de Janeiro), important political episodes (such as the falling of the Berlin Wall, or the changing of political parties in a government, such as Margaret Thatcher taking power in the United Kingdom), normative dispositions with the force of soft or hard laws (such as the signing of an international treaty, joining an international organization or declaring adherence to a cause, or the formation of political coalitions), and the publication of a work or technical report by an international organization (like *Development as Freedom* by Amartya Sen or the World Bank’s *World Development Report*).

A comparison may be structured in many different ways. A study may be designed for the systematic observation of processes and the adoption of similar policy models by different governments, such as municipal or national governments, or even distinct regions. Another way of comparing consists of the analysis of different policies or instruments of public action in the same sector, such as health care, the environment, or education. Now we will look at a few examples of how researchers have designed their studies and the results they have obtained.

One way to compare policy transfers is analyzing the adoption of similar models in cities within different countries. There are a large number of “best practices” in public management these days which circulate internationally, such as PB, bus rapid transit (BRT) systems, and CCT programs, among others. In terms of the diffusion of BRT policies, an illustrative example is the study by Mejía-Dugand et al. (2013) which collected information about 30 countries which have adopted the BRT system in Latin America in order to create a map of the adoption process, its temporality, and the transformations of the model. The authors argue that the diffusion process for the BRT has been associated with an incremental movement in the accumulation of its instruments. Initiated in Latin America based on the precursor experience of Lima, Peru in 1972, the system was then perfected in a concrete manner in Curitiba, Brazil, culminating in Bogotá, Colombia, which improved the model even further (Mejía-Dugand et al. 2013: 86). This study states that the diffusion of BRT in Latin America has followed a flow of non-hierarchic connections, in which the “element (city) is affected/influenced and at the same time each element affects/influences any other element” (Mejía-Dugand et al. 2013: 85).

The authors identify waves of diffusion, occurring before Curitiba, between 1972 and 1997, and the period that followed the implementation of the Bogotá system in 2011, when the volume of adoptions became greater. The transfer process and the improvement of the system made the model more adaptable to other contexts, and reports of the positive results of this policy stimulated its diffusion in the region (Mejía-Dugand et al. 2013: 88). This transnational

flow is demonstrated in the text when the authors affirm that “politicians and planners from Curitiba visited the Peruvian system” and soon after adapted this model. However, it was only with the Bogotá experience that this system achieved international recognition, which made room for the creation of technology exchanges associated with BRT. “Latin American cities seem to have created a network that has facilitated the propagation of innovative solutions that share a common view, while focusing on their own resources” (Mejía-Dugand et al. 2013: 86).

A comparison can be restricted to cities within the same country. Astrid Wood (2015), for example, has also analyzed the adoption of BRT policies. However, in her study she just considers local South African governments. She reconstructs the trajectory of the adoption of BRT, focusing on Johannesburg and Cape Town, including their transnational connections and the contacts that these cities had with Curitiba and Bogotá. By tracing the process of this innovation, the author concludes that even though BRT arrived in the country in 2006, the implementation process was gradual and slow. There had already been previous experiences with similar models in the 1970s. She observes the mechanisms that operated at different times, such as learning from the Curitiba model (just after apartheid ended in the 1990s), which was considered a less expensive and more flexible solution than other options available (Wood 2015: 576). Actors in the South African transport sector traveled to this Brazilian city frequently to learn about their experience. During the following decade, however, Bogotá became a reference for these South African cities. The adoption of the BRT model was repeatedly delayed as a result of political and bureaucratic interference, as well as reasons related to the learning process. Given the pressure to construct infrastructure for the 2010 World Cup, several South African cities then finally organized themselves to implement BRT.

The analysis of the BRT transnational process and the local implementation from a historical perspective enables Wood to make an important statement about the field of policy transfer, which is that policy diffusion occurs slowly nowadays. Wood (2015: 578) demonstrates that there exist “multiple temporalities of BRT learning, framing contemporary adoption practices with the protracted and idiomatic nature of policy circulation processes”. This statement runs counter to a recent emphasis in the literature on the rapid nature of international policy spread, which was popularized through the expression “fast policy” by Peck and Theodore (2015). In fact, the historical comparison conducted by Wood makes it possible to argue that while “local decisions regarding circulated ideas can be delayed by a lack of funding, postponed because of politicking, or become stuck in bureaucracy, learning is ongoing” (Wood 2015: 578).

Also in Latin America, Cecilia Osorio Gonnet (2019) compares CCT policies in different countries, namely Chile and Ecuador, to try to understand the causal mechanisms that facilitate the model adoption process. Today CCTs exist in 18 of the 20 countries in Latin America and are characterized by transferring income to poor or extremely poor families, with the beneficiaries needing to meet conditions to have access to this income which are usually related to health and education services. Her methodology uses the technique of different systems and these cases are heterogeneous in terms of poverty, inequality, GDP, and ideology, but both of these countries adopted a similar policy, Chile Solidarity in Chile and the Human Development Bonus in Ecuador.

This study observes, on one hand, epistemic communities and their relationships with international organizations and the technical capacities of state institutions and, on the other, it tests coercion, learning, and emulation mechanisms. In her comparison, the author identifies that there is an important relationship between the presence of epistemic communities and state capacities, which are elements that can facilitate the adoption of CCT programs.

In Chile, this process occurred through learning by domestic actors, in addition to a limited role played by external forces such as the World Bank, which is one of the great promoters of CCT policies and was also present in the Chilean case (Osorio Gonnet 2019: 391). In Ecuador there was a coercion mechanism exercised by multilateral banks in the first phases of the policy implementation through loans and technical cooperation (Osorio Gonnet 2019: 393). However, during the second implementation phase, domestic actors assumed the leadership of the program.

Cases can also be defined as policy sectors such as: social assistance, health care, education, retirement, immigration, taxation, etc. Patrick Hassenteufel et al. (2017) deal with the case of health care policy in Europe, comparing the United Kingdom to France and Germany. This study is devoted to Evidence-Based Medicine (EBM), whose origin dates back to the United States in the 1970s, and it has been disseminated on an international scale. The authors insist, in particular, on the analytical potential of the policy transfer dimension, which involves the observation of discourses, power relationships between actors, and the adaptation of models (Hassenteufel et al. 2017: 82). In process tracing the adoption of EBM in the United Kingdom in 1999, they observe that this country produced important innovations in terms of management and indicators such as the National Institute of Clinical Excellence and created institutions to encourage its export to other countries. The British model is characterized by a bureaucracy based on evidence which uses cost–benefit evaluations in a systematic manner, which are very structured in protocols and standards, and which are also open to non-governmental actors (Hassenteufel et al. 2017: 83).

France has sought to mimic the United Kingdom, but not in an explicit manner as in Germany, in which the references to the adoption of the British model are obvious. In examining the German experience, the authors insist on the action of physician interest groups, which play an important role in the country's health care policy. These actors influence the policy transfer process, with one of the effects being the abandonment of the cost–benefit dimension in the evaluation of medication and medical interventions. In Germany, the result of this translation has been the exclusion of a strong bureaucracy based on evidence as well as a systematic manner of using economic knowledge in policy development, so that the model in reality has only been partially adopted from the original British model. In turn, in France the authors point to a movement of incremental translation where there have been progressive advances in the formation of a consensus among health care economists in the administration of this sector. These actors have inserted proposals in health care institutions which are contrary to those of the medical community, creating internal and external pressure. The result of this has been the adoption of “a centralized evidence-based bureaucracy using the same cost-benefit assessment tools” as the British model (Hassenteufel et al. 2017: 89).

In our study of the global diffusion of PB, the comparative strategy used techniques of similar systems design to understand why the internationalization of the model adopted by the city of Porto Alegre had greater prominence than those of Belo Horizonte and Recife (Porto de Oliveira 2017). PB is *grosso modo* a participatory institution, through which citizens can decide directly where they want public spending to be invested, in terms of local public policies. It was developed in Porto Alegre in 1989 by the left-wing Workers Party, then rapidly adopted by other municipalities in Brazil in the 1990s and reached almost 3,000 cases all over the world, from New York to Cheng Du (China), passing through Maputo (Mozambique) and Paris (France). The three Brazilian cities under analysis developed PB experiences, with some degree of internationalization, with Porto Alegre being the pioneer. All are state capitals in

different regions of Brazil, with Porto Alegre being the capital of Rio Grande do Sul, Belo Horizonte the capital of Minas Gerais, and Recife the capital of Pernambuco.

The similar elements which were shared by these cities were the following: the international action of these municipalities through the organization of events, fundraising through international organizations, the creation of institutions to promote “city diplomacy” (such as secretariats and coordinating bodies in terms of international relations), the construction of transnational networks, and receiving awards from foreign institutions, as well as the building of a distinctive “brand” related to the PB model. After applying the MSSD it was possible to determine that one of the elements that has made a difference in the case of Porto Alegre has been the engagement of high-level political officials in the process of internationalizing their experience. This signifies that successive mayors in Porto Alegre during the 1990s led strategies to internationalize PB, which was an important element in the political agenda of these administrations. The other cases of internationalization were conducted by less powerful political figures such as secretaries, their advisors and other bureaucrats (Porto de Oliveira 2017: 88) and the internationalization of PB was not a priority on the political agenda.

The transfer process can involve a vast array of agents, instruments, and arenas. A portion of the studies related to this area are dedicated to analyzing the role of these elements in the diffusion of policies. We will see how this dimension is analyzed in the next section.

6. COMPARING TRANSFER AGENTS, INSTRUMENTS, AND ARENAS

This field has attached great importance to the international dimension of transfers and many studies are dedicated to understanding the agents involved in transfer processes, the instruments that circulate from one place to another, and the arenas in which transnational interactions occur. These intervening variables are incorporated in these studies as explanatory elements of the transfer process or as objects of analysis *per se*. In the following sub-sections we will see examples of comparisons involving these elements.

6.1 Transfer Agents

Transfer agents can be individual, collective, governmental, inter-governmental, non-governmental, domestic or international. Studies have been interested in the role of humans as policy entrepreneurs or “ambassadors”; think tanks in the promotion of sectoral policies, such as the environment or health care; and international organizations such as the European Union, the World Bank, and agencies of the United Nations system. In order to study PB, we analyzed the role of so-called “ambassadors of participation”, who are true transnational militants of social participation on public policy making and dedicate a good portion of their time and energy to promoting participatory institutions, which is often a life mission. The technique of analysis was to follow a group of individuals that were active in the promotion of PB in different countries. The observation of these actors allowed us to discover characteristics of a new type of agent operating on policy transfers, especially from the South, and develop a concept of “policy ambassadors” to better understand the phenomenon. Ambassadors of participation have political authority, expertise, and technical experience (Porto de Oliveira 2017: 48). The systematic observation of the international circulation of these “ambassadors

of participation” in various countries in Latin America, Europe, and Africa, and in various governmental, non-governmental, and academic institutions made it possible in our study to affirm that their actions were a *sine qua non* condition for PB’s inclusion on the international agenda. In addition, it has been possible to identify distinct political projects behind the devices these agents were carrying out transnationally. Finally, this analysis shows that these actors support PB no matter which institution they are affiliated with, whether they are universities, city governments, or even international organizations. Without the actions of these actors, PB probably would not have become globalized to the extent that it has.

Stella Ladi (2005), in turn, dedicates herself to understanding the role of three research institutes in policy transfers in the European Union in terms of labor (International Dialogues Foundation), the environment (Understandingbus), and social control (Paremvassi). In her analysis, the author insists on the importance of think tanks as agents for the diffusion of ideas and policies. In her comparison, the similarities and differences between the actions of each of these research institutes, the European Union, and the national governments involved are observed during a series of steps in the policy transfer process. More importantly, she compares the roles of these actors, their specific actions, and their passivity, when they do not exercise their influence. The phases defined by Ladi correspond to the moments that precede the decision-making process in the implementation of the policy, in the following order: (1) recognition of the problem; (2) seeking solutions; (3) contacting research institutes; (4) the emerge of an information feeder network; (5) the recognition, acceptance, and emergence of a transfer network; (6) mobilizing and recognizing ideas and the elite; (7) interactions between actors (politicians and research institutes); (8) evaluation of the best alternative to be implemented; (9) the making of a decision; and (10) its implementation (Ladi 2005: 158–60).

Ladi finds that the European Union exercises an important role and can influence transfer ideas and priorities (Ladi 2005: 156) and that think tanks are part of this structure. At certain points in the transfer process, some think tanks do not play a significant role. Three findings in this comparison deserve attention based on their potential in terms of general theory. The role of think tanks in the transfer process varies according to the political and administrative system and the maturity of the civil society in which this policy is being adopted (Ladi 2005: 162). In terms of the roles that think tanks play in each of these cases, there is an interdependent relationship between them and the European Union and national governments. On the one hand, these governments need think tanks to produce information, and, on the other, these research institutes depend on governments and the European Union for their financing. Finally, the motivation behind the work of these think tanks is associated with promoting the cause that they work for, not with the financial resources themselves (Ladi 2005: 163).

International organizations are among the most popular actors in policy transfer studies, especially in their role in inducing or using coercion to convince governments to adopt these policies. International organizations’ induction mechanism can be executed through conditionality clauses, technical assistance, the production of information, and even the design of programs. In this way, we can observe the direct influence of international organizations in the domestic sphere, as well as their role in the global diffusion of models and ideas, often in the form of “best practices”. The World Bank, the International Monetary Fund (IMF), and the Organisation for Economic Co-operation and Development (OECD) are some examples of active agents in the diffusion of policies.

The OECD is an important subject in these studies, due to the centrality of the production and diffusion of ideas and knowledge in its actions. The OECD uses transfers as a mechanism

to help state reform processes with the objective of supporting economic development. Leslie Pal (2012) has conducted an extensive investigation about this organization and its role in the diffusion of new public management (NPM) instruments. His historical analysis traces the institutional evolution of the OECD since its beginnings with the Marshall Plan, which shows the way in which this institution has specialized in being a global agent dedicated to “knowledge creation and dissemination” (Pal 2012: 50) on public administration, especially after the fall of the Berlin Wall, when a large number of ex-Soviet bloc countries needed to make the transition to liberal democracies and market economies. In this case study, Pal presents a perspective that is innovative for this field because it analyzes the policy transfer construction dimension from the internal point of view of international organizations, and not from the classical perspective, which is about the actions of organizations towards other countries. In the following section, we will deal with the European Union and other countries, presenting how the authors have structured their comparisons with a more direct focus on the transfer objects.

6.2 Transfer Instruments

The European Union has created its own research area, Europeanization studies, paving the way for a dense empirical field of policy transfer comparative analyses. The objects of comparisons are not centered only on countries, but also on transfer instruments. In fact the part of the regional integration process that is associated with the harmonization of public policy norms, procedures, and instruments among these countries has merged studies of transfer and Europeanization (Saurugger and Surel 2006). This has led many studies of Europeanization to also deal with policy transfers, without necessarily making specific mention of the latter term or using this field’s references. In effect, as Halpern and Le Galès (2011: 47) note, the “evolution of European public policy instruments largely results from the diffusion, the transfer and the rearrangement of pre-existing elements, which raises specific questions of sedimentation, inertia and change, possible contradictions and growing demands for coordination”.

Halpern and Le Galès (2011) focus their research on the systematic analysis of the instrumentation of European Union policies, aspiring to explain the governance process on a regional level. To the authors, instruments are “a device that is both technical and social, that organises specific social relations between the state and those it is addressed to, according to the representations and meanings that it carries” (Halpern and Le Galès 2011: 61). In their study, they compared 16 instruments of environmental (53) and urban (33) policies between 1972 and 2006, and observed that the development of these sectors is associated with the produced instruments. In relation to the instruments in the environmental sector, the actors identify that there has been little capacity for innovation in the European Union, given that two-thirds of the instruments have been transferred in general from the international field, through accords, treaties or member states (2011: 54). In the case of urban policy, instruments and innovations produced by the European Union are prevalent, but they have not necessarily been developed by this sector. It should be observed that the origin of these instruments and their nature helps us understand the governance process in European regional integration and the similarities and differences among these sectors.

These instruments can be transferred in blocks, or in other words, as solution packages for problems of a public nature. This is the case of state reforms and the global diffusion of NPM instruments, as we have observed in the work of Leslie Pal presented above. Guy Peters (1997), in a pioneering study of the policy transfer area, compares Eastern European

countries in the adoption of types of policies which make up three different large areas: market reforms, participation, and deregulation.⁷ In his comparison, the author considers administrative reforms and emphasizes the internal political dimension of states, including independent variables that influence the process such as: (1) the ideological orientation of the party (right or left) that is governing (Peters 1997: 76); (2) the power of public unions (Peters 1997: 77); (3) the socio-economic context, considering the size of the public debt (Peters 1997: 77); and (4) the administrative culture (Peters 1997: 78). In observing the presence of each element in the adoption of policies in distinct countries, the author reaches the conclusion that culture plays an important role, while the other variables, such as the political party in control of the executive branch and the size of the deficit have less influence. This finding is unexpected for the literature at this time, which tended to favor arguments that associate the adoption of similar policies with the geographic proximity of countries and other forces.

Besides instruments, it is important to compare spaces where agents interact, or in other words, where they exchange experiences, legitimize models, and argue their points of view. These spaces are called transfer arenas, and we will see how they can be incorporated in comparative analysis.

6.3 Transfer Arenas

Transnational meetings such as summits promoted by countries or international organizations (COP 21, UN-Habitat, the G20 Trade Ministers Meeting), thematic forums (Global Child Nutrition Forum, Metropolis, World Economic Forum), and training workshops, among others have become recurrent mechanisms for the exchange of ideas and the building of global agendas. In fact, within the analysis of domestic processes there are important areas such as “courts”, legislatures, forums of civil society, party conventions and various national meetings which constitute the arenas in which public policy is negotiated and defined. These days, the transnational meetings that we have mentioned correspond to the international face of similar spaces on the domestic level. Diane Stone calls some of these new spaces “global agora” in reference to the squares in which politics were discussed in Ancient Greece. In her words, such a place is a “global public space of fluid, dynamic, and intermeshed relations of politics, markets, culture and society” which is characterized by “the interactions of its actors – that is, multiple publics and plural institutions” (Stone 2008: 21).

Analyses can use “zoom in and zoom out” strategies to understand elements of these spaces such as: who participates, the power relationships, what is argued, what are the participants’ strategies, what discourses are proffered, which models are legitimized or delegitimized, and the construction of the agenda. In a recent study, Aykut, Foyer, and Morena (2017) use an ethnographic study to understand the globalization of the agenda to fight climate change, in which they use team field research of the mega transnational event of the 21st Conference of the Parties Climate Convention (COP 21), which occurred in Paris in 2015. In their analysis, the authors pay attention to the changes in the meanings of the discourses related to climate over time as well as the solutions proposed and the action strategies for global governance in this sector.

In studying the process of the international diffusion of PB in our work, we also sought to understand transnational transfer arenas. Process tracing led to the identification of city meetings as having been crucial to bringing people together, and serving as a trigger for cooperation projects with the object of transferring PB, legitimizing this practice on an international scale,

and continually increasing the number of its potential adopters. The arenas were analyzed through documents in terms of previous field research, and participative observations in relation to what occurred during the investigation. In particular, we applied a survey at the Africities Forum (2012), which is the largest meeting dealing with city issues in Africa, and takes place every three years. Questionnaires were given to 84 participants in the sessions that dealt with PB. A comparison of the responses to the survey made it possible to identify the profiles of the participants, to estimate the number of people who were coming in contact with this experience for the first time, and to analyze how each respondent perceived PB in the region of sub-Saharan Africa, or in other words, it was possible to get a deeper perspective of what Africans thought of PB at that event. In addition, the comparative observations of these sessions were important in identifying which experiences received the greatest amount of attention, receiving more space at tables, and which had less visibility, which helped identify the legitimization of these practices. Finally, by observing the schedule of the events from a historical perspective, it was possible to understand which cities received the greatest amount of attention at different times, and who were the protagonists of the event – those individuals who always appeared on the schedule.

7. CONCLUSION

The analysis of international policy transfers is fundamental to understand contemporary public policies. In order to access such phenomena it is necessary to move beyond conventional comparative public policy methods and overcome “methodological nationalism”. Scholars working with policy transfers frequently include the transnational dimension in their analysis. The association of the observation of transnational features of policy transfers to traditional comparative methods as MSSD and MDSD is a solution, to deal with the issue of interdependence, as well as a strategy to capture the movement and transformation of policy instruments. Researches have been focusing on middle- and small-N comparisons, as well as using case studies as a comparison. Process tracing is a common technique for studies focusing on the international transfer process. There are different possibilities for analyzing processes, which includes the comparison of the adoption of the same policy in different cities or countries. By analyzing international policy transfer processes, authors can compare different elements along the transnational movement, such as agents, instruments, and arenas.

If there has been much comparative work done to analyze international policy transfer processes, there is still little about the impact of transfer on policy outcomes. It is well known that international forces have an influence on domestic policies. However, there is much to discover about how transnational variables affect policy “success” and “failure”. This is still a frontier to be broken in the literature and Diane Stone (2017) has advanced this topic by addressing the issues of uninformed, uncompleted, inappropriate transfers and negative lesson-drawing, but there is still much to be done about how to compare these cases, as well as negative transfer cases.

Even if there is a wide range of topics in the international policy transfer area of research that still needs to be covered, the field is building an identity. This unity of the area has showed common comparative strategies to analyze public policy transfers, offering a solution to overcome the Galton problem, as well as opening an opportunity for methodological innovation on the broader area of comparative public policy. In a context of globalization

and resistance – where progressive and conservative political worldviews of state institutions and public administration flow rapidly across countries – there is an imperative to understand public policies in movement in comparative fashion and there is still need to formalize and sophisticate existing techniques.

NOTES

1. This was one of the most popular areas at the Third Annual International Conference on Public Policy, organized by the International Public Policy Association in Singapore, and it was also a central theme at the International Conference on Policy Diffusion and Development Cooperation, organized by the Federal University of São Paulo, Brazil, which involved roughly 270 registered participants.
2. In a bibliometric study, Graham, Shipan, and Volden (2013) found more than 100 terms to describe the phenomenon of policy transfer.
3. See Clarke et al. (2015) and Peck (2011) in relation to this.
4. See Porto de Oliveira and Pimenta de Faria (2017) in this regard.
5. The literature on policy diffusion, especially in the United States, uses advanced quantitative techniques to make comparisons with large-N samples to understand this phenomenon. Given that this chapter is dedicated to qualitative analyses, we will not treat this dimension here. To find out more about quantitative techniques and the related literature see Graham, Shipan, and Volden (2013).
6. This perspective is also shared by studies in the field of geography which use the notion of “policy mobility” (Clarke et al. 2015).
7. The author classifies the variation of each type of reform through subtypes: (1) Market Reform (Agency Model, Pay-for-Performance, Internal Markets, Performance Contracts, Accrual Accounting, Program Reviews, One-Stop Shops); (2) Participation (Quality Management, Decentralization, Citizen Charters); (3) Deregulation (Bulk Budgeting, Purchasing Deregulation, Personnel Deregulation) (Peters 1997: 74).

REFERENCES

- Aykut, S. C., Foyer, J., and Morena, E. (2017). *Globalising the Climate: COP21 and the Climatisation of Global Debates*. New York: Routledge.
- Braun, D. and Gilardi, F. (2006). Taking “Galton’s problem” seriously: towards a theory of policy diffusion. *Journal of Theoretical Politics* 18 (3): 298–322.
- Bruno, I., Jacquot, S., and Mandin, L. (2006). Europeanization through its instrumentation: benchmarking, mainstreaming and the open method of co-ordination ... toolbox or Pandora’s box? *Journal of European Public Policy* 13 (4): 519–36.
- Clarke, J., Bainton, D., Lendvai, N., and Stubbs, P. (2015). *Making Policy Move: Towards a Politics of Translation and Assemblage*. Bristol: Policy Press.
- Dezalay, Y. and Garth, B. G. (2002). *La mondialisation des guerres de palais: La restructuration du pouvoir d’État en Amérique Latine, entre noblesse du droit et “chicago boys”*. Saint-Amand-Montrond: Éditions du Seuil.
- Dolowitz, D. P. and Marsh, D. (2000). Learning from abroad: the role of policy transfer in contemporary policy making. *Governance* 13 (1): 5–24.
- Dunlop, C. A. (2009). Policy transfer as learning: capturing variation in what decision-makers learn from epistemic communities. *Policy Studies* 30 (3): 289–311.
- George, A. L. and Bennett, A. (2005). *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- Gerring, J. (2004). What is a case study and what is it good for? *American Political Science Review* 98 (2): 341–54.

- Goldthorpe, J. H. (1997). Current issues in comparative macrosociology: a debate on methodological issues. *Comparative Social Research* 16: 1–26.
- Graham, E. R., Shipan, C. R., and Volden, C. (2013). The diffusion of policy diffusion research in political science. *British Journal of Political Science* 43 (3): 673–701.
- Hadjiisky, M., Pal, L. A., and Walker, C. (Eds.) (2017). *Public Policy Transfer: Micro-Dynamics and Macro-Effects*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Halpern, C. and Le Galès, P. (2011). No autonomous public policy without ad hoc instruments. *Revue Française de Science Politique* 61 (1): 43–67.
- Hassenteufel, P. (2005). De la comparaison internationale à la comparaison transnationale: les déplacements de la construction d'objets comparatifs en matière de politiques publiques. *Revue Française de Science Politique* 55 (1): 113–32.
- Hassenteufel, P., Benamouzig, D., Minonzio, J., and Robelet, M. (2017). Policy diffusion and translation: the case of evidence-based health agencies in Europe. *Novos Estudos – CEBRAP* 36 (1): 77–96.
- Hassenteufel, P. and de Maillard, J. (2013). Convergence, transferts et traduction: les apports de la comparaison transnationale. *Gouvernement et action publique* 3 (3): 377–93.
- Howlett, M., Ramesh, M., and Saguin, K. (2018). Diffusion of CCTs from Latin America to Asia: the Philippine 4Ps case. *Revista de Administração Pública* 52 (2): 264–84.
- Huntington, S. P. (1993). *The Third Wave: Democratization in the Late Twentieth Century*. Norman, OK: University of Oklahoma Press.
- Ladi, S. (2005). *Globalization, Policy Transfer and Policy Research Institutes*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Landman, T. (2007). *Issues and Methods in Comparative Politics: An Introduction* (3rd edition). New York: Routledge.
- Mejía-Dugand, S., Hjelm, O., Baas, L., and Ríos, R. A. (2013). Lessons from the spread of bus rapid transit in Latin America. *Journal of Cleaner Production* 50: 82–90.
- Osorio Gonnet, C. (2019). A comparative analysis of the adoption of conditional cash transfers programs in Latin America. *Journal of Comparative Policy Analysis: Research and Practice* 21 (4): 385–401.
- Pal, L. A. (2012). *Frontiers of Governance: The OECD and Global Public Management Reform*. Basingstoke: Palgrave Macmillan.
- Peck, J. (2011). Geographies of policy: from transfer-diffusion to mobility-mutation. *Progress in Human Geography* 35 (6): 773–97.
- Peck, J. and Theodore, N. (2015). *Fast Policy: Experimental Statecraft at the Thresholds of Neoliberalism*. Minneapolis: University of Minnesota Press.
- Peters, B. G. (1997). Policy transfers between governments: the case of administrative reforms. *West European Politics* 20 (4): 71–88.
- Porto de Oliveira, O. (2017). *International Policy Diffusion and Participatory Budgeting: Ambassadors of Participation, International Organizations and Transnational Networks*. Cham: Palgrave Macmillan.
- Porto de Oliveira, O. and Pal, L. A. (2018). New frontiers and directions in policy transfer, diffusion and circulation research: agents, spaces, resistance, and translations. *Revista de Administração Pública* 52 (2): 199–220.
- Porto de Oliveira, O. and Pimenta de Faria, C. A. (2017). Research traditions and the state of the discipline in Brazil. *Novos Estudos – CEBRAP* 36 (1): 13–34.
- Robinson, J. (2018). Policy mobilities as comparison: urbanization processes, repeated instances, topologies. *Revista de Administração Pública* 52 (2): 221–43.
- Rose, R. (1991). What is lesson-drawing? *Journal of Public Policy* 11 (1): 3–20.
- Saburin, E. and Grisa, C. (Eds.) (2018). *A difusão de políticas brasileiras para a agricultura familiar na América Latina e Caribe*. Porto Alegre: Escritos.
- Saurugger, S. and Surel, Y. (2006). L'européanisation comme processus de transfert de politique publique. *Revue internationale de politique comparée* 13 (2): 179–211.
- Simmons, B., Dobbin, F., and Garrett, G. (Eds.) (2008). *The Global Diffusion of Markets and Democracy*. Cambridge: Cambridge University Press.
- Stone, D. (2001). Think tanks, global lesson-drawing and networking social policy ideas. *Global Social Policy: An Interdisciplinary Journal of Public Policy and Social Development* 1 (3): 338–60.
- Stone, D. (2008). Global public policy, transnational policy communities, and their networks. *Policy Studies Journal* 36 (1): 19–38.

- Stone, D. (2017). Understanding the transfer of policy failure: bricolage, experimentalism and translation. *Policy & Politics* 45 (1): 55–70.
- Weyland, K. (2007). *Bounded Rationality and Policy Diffusion*. Princeton, NJ: Princeton University Press.
- Wimmer, A. and Glick Schiller, N. (2002). Methodological nationalism and beyond: nation-state building, migration and the social sciences. *Global Networks* 2 (4): 301–34.
- Wood, A. (2015). Multiple temporalities of policy circulation: gradual, repetitive and delayed processes of BRT adoption in South African cities. *International Journal of Urban and Regional Research* 39 (3): 568–80.
- Woods, N. (2006). *The Globalizers: The IMF, the World Bank, and Their Borrowers*. Ithaca, NY: Cornell University Press.

PART III

MEASUREMENT AND EXPERIMENTAL METHODS

9. Using experiments in comparative policy analysis: from policy evaluation to the policy process

Peter John

1. INTRODUCTION

Since the mid-1990s there has been a remarkable renaissance in the use of experiments to answer a diverse range of social science research questions. Examples include the evaluation of development programmes (Fearon, Humphreys, and Weinstein 2009), ways to increase voter turnout (Gerber and Green 2000), anti-corruption measures (Chong et al. 2015), crime reduction (Sherman and Weisburd 1995), and detecting discrimination (Butler and Broockman 2011), to name only a few applications. These research projects share the use of random allocation or its equivalent to determine whether an intervention or process has the intended or hypothesized effect and they assess by how much is the impact in point estimates. The comparison of outcomes in a group that gets an intervention with one that does not generates a valid causal inference (Gerber and Green 2012), which is a very powerful way to add to knowledge as the technique may be deployed to answer authoritatively a precise research question (Glennerster and Takavarasha 2013; Torgerson and Torgerson 2008), even when taking into account the limitations of experiments (Cartwright and Hardie 2012; Deaton and Cartwright 2018). As many commentators have pointed out (e.g. John 2017), this rediscovery is something of a natural correction, which returns social science to the concerns and methods of its founders. Experimentation was the core part of statistical theory as it developed in the early twentieth century; even regression analysis, the workhorse of the analysis of observational designs, originates in experiments (Angrist and Pischke 2008). The 1920s saw the development of field experiments in health, education, and political participation. But, outside health procedures and drug tests, field experiments languished until their enthusiastic rediscovery in the second half of the twentieth century, starting with the work of Campbell (Campbell 1957), then with their wholesale adoption from the 1990s. Along with the use of randomized interventions, there has been a development in the statistical theory of causal inference, which has underpinned the development of what are called natural experiments (Dunning 2012), as well as greater attention to quasi-experiments (Shadish, Cook, and Campbell 2002; Shadish, Cook, and Leviton 1991) and identified designs (Angrist and Pischke 2008).

If experiments have been frequently used to answer practical questions in the evaluation of public policies, they have hardly appeared at all in the studies of the policy process, which is the more political science end of public policy, and represent the core set of issues that many public policy scholars examine today to understand and explain decision making, especially in comparative context. This might be thought to be a function of the nature of experiments in that they are attuned to precise evaluation questions rather than assessing complex decision processes. The policy evaluation question of whether a welfare-to-work policy actually works

is done by randomizing the new intervention to one group of welfare beneficiaries and not another; based on the findings an agency can decide whether to introduce the policy. But this is a long way from understanding decision making, such as whether policy punctuations are caused by friction (Jones and Baumgartner 2005), for example, which might be hard to test with an experimental method because it may not be possible to locate the random allocation or create an intervention that might test the hypothesis effectively. NP The theme of this chapter is that such pessimism is not fully justified and that there remains an opportunity to use experiments more extensively in studies of the policy process and in comparative public policy more generally. As well as field trials, there may be an opportunity to use natural experiments to test for theories of the policy process. To make this argument the chapter shows how political scientists have recently directed their attention to decision-making processes and institutions with great effect, creating a set of second-generation experiments different from the first-generation ones mainly done on political behaviour. Even more importantly, and one for public policy scholars to take notice of, there has been a parallel development in the field of public management, which is a cognate field to public policy studies. With methods in public management previously being confined to survey methods, observational aggregate data, and case studies, the 2010s have seen a veritable explosion of experimental studies which seek to examine decision making within and across public bureaucracies, such that this sub-discipline has changed dramatically in terms of the questions that are asked and the manner in which they are evaluated (James, Jilke, and Van Ryzin 2017). That the public management sub-field has moved so fast is both a challenge and a prompt to scholars of public policy to adopt new methods, to answer old and new research questions with experiments. Fortunately, there are some signs of experimental methods recently appearing in public policy, mainly the use of new methods for observational data and the analysis of some natural experiments. But much more needs to be done, especially more use of randomized controlled trials; and this chapter is designed as an introduction to researchers and students to give them necessary knowledge about how it could be done.

The next section starts with a basic introduction to experiments and notes the history of their use in answering public policy questions in evaluation. It then reviews the expansion of experiments in political science, especially elite experiments, then their appearance in public management, with the idea that these designs are in fact answering questions public policy scholars typically pose, such as assessing biases in decisions made by politicians and bureaucrats.

2. WHAT ARE EXPERIMENTS?¹

An experiment occurs when human beings manipulate the world to understand causal relationships. It is a common means by which scientists acquire knowledge. The secret is to have precise control over the different elements of an intervention and to measure carefully what happens during the experiment. It is this degree of control over the external world that is the big attraction of experiments, which explains why researchers from other disciplines seek to emulate the method. By intervening as well as observing an experiment unlocks knowledge in ways that cannot be fully achieved by observation alone.

In the social sciences and in government, the researcher or policy maker looks for or creates some random or chance variation that ensures that the difference in outcomes is only associated with an intervention or policy compared to the status quo. In what is called a natural

experiment, accidental differences between populations or areas that have been created by government or by nature, such as the impact of boundaries or cut-offs for eligibility in a policy programme, may be thought of as if they were random (Dunning 2012). Sometimes randomness happens directly, such as in a government decision or institutional rule that can be evaluated like an experiment.

More often, researchers and policy makers create the random variation themselves. This method is sometimes called artificial randomization, but is more commonly known as the randomized controlled trial, where individuals or communities or other units are randomly allocated to two or more groups. In this procedure, one or more groups get an intervention while another group, the control, does not. A comparison can be made between the outcomes in these groups to see if the intervention (also commonly called the treatment) made an impact. If there are at least two comparison groups and one group receives an intervention while another does not – or gets a different intervention – and if assignment to these groups is random, then it is possible to make an inference that the difference in outcomes between the groups – other than random variation – has only been caused by the intervention.

This kind of method is sometimes called a field experiment, which is done in the community at large or within an organization. It may be contrasted with a laboratory experiment done in controlled settings (excepting the hybrid form called ‘lab in the field’ where a laboratory experiment is done in the field). More generally, people use the term ‘trial’ or the ubiquitous acronym, RCT. For example, if a public agency wants to test whether giving training gives the unemployed more of a chance to find employment it can find a group of welfare recipients to try out an intervention; the agency can then randomly allocate the training to one group and leave the other alone or just provide a normal level of service. The agency can then find out if those who got the training were more likely to find employment than those who did not.

3. CAUSAL INFERENCE

One of the most attractive features of a trial is that it can offer a clear test of a hypothesis, usually resulting in either a yes or no to the research question. For policy makers, a well-designed trial can confirm whether an intervention worked or not. Moreover, if enough trials are done, it is possible to conclude that it can be generalized across places and time periods. It is the claim to answer questions about causation that is the main appeal of a trial and explains why it has become so popular as a method in recent years.

It might sound surprising that social science, which started using advanced statistical methods in earnest from the 1940s, should still want better leverage on causal questions. But it is very hard to make a causal inference in social research, even when there is a very strong theory about the determinants of a particular outcome and good measurements of what might cause it. This weakness occurs because it is not possible to rule out with certainty that factors other than the hypothesized one caused the outcome of interest or to confirm the direction of causation in a correlation. In social science as it has been practised over the last 50 years or so, it is common to observe the simultaneous occurrence of an influencing and influenced variable and to measure its association.

Researchers who analyse observational data have been aware of the problem of establishing causation for a long time. They have developed a number of strategies to overcome it. One is to consider the possible correlates between the outcome and the intervention, collect data on each

one and then control for the association between the hypothesized variable and the outcome of interest. This strategy is achieved by multiple regression, whereby several variables, including the one of interest, are treated as independent causes of the outcome. If the association between the independent variable of interest and the outcome still remains as statistically significant after this procedure has been followed, it can be more reasonably concluded that there is a causal relationship. This inference is thought to be valid because other causal pathways have been allowed for and the item of interest is still a factor determining the outcome. This strategy becomes more convincing if the researcher has considered in advance all possible causes of an outcome from theory and existing empirical work, measured and collected data on these alternative explanations, and introduced the causal propositions as independent variables in the statistical model.

Yet it may be possible that the researcher misses an alternative explanation or cannot accurately measure what causes the outcome, making the causal inference hard to support even when control variables have been used. It is often said that the relationship is confounded or that there are confounders in play, that is, factors that are associated with both the intervention and the outcome being evaluated. The researcher can only guess what they are and cannot effectively control for them. There may also be other unobserved processes at work. Another problem is that the outcomes for individuals or units will vary over time. Their outcomes might get worse or better in ways that follow a natural cycle, such as finding a job, becoming healthier or the opposite. In this case, all the programme or intervention picks up is what would have happened anyway – the counterfactual, which is the state of affairs that would have occurred in absence of the programme. In all these circumstances, it is important to have a method that rules out determinants of an outcome other than the intervention. Randomization generates the counterfactual of what would have happened without the intervention (Gerber and Green 2012), such as between unemployed people some of whom received a job training programme while other randomly selected people did not.

4. THE HISTORY OF EXPERIMENTS IN PUBLIC POLICY

Experimentation has been around as a research method for a long time. The idea of casting lots in order to allocate people to treatments occurred in the seventeenth century (Michalopoulos 2005: 11). The first practical examples of having a comparison group of people probably goes back to James Lind's 1747 experiments with citrus fruit to treat sailors for scurvy. The practice of randomly allocating subjects to test an intervention originates in the nineteenth century, which was the start of experimental psychology. Also important was the common practice of alternating medical treatments, which had started in the 1860s. This practice is the non-random allocation of treatments to some patients and not to others, which of course does not meet the demands of randomization, but can rightly be seen as the precursor to the trial and where randomization can be added in as a procedure to select the patients as later researchers did.

The scientific study of field experiments properly goes back to the work of Ronald Fisher, who worked in agriculture. His work appeared in a number of papers going back to the 1920s (Fisher 1926), which were summarized in his famous book, *The Design of Experiments* (1935). While agricultural researchers had given careful attention to testing and measurement beforehand, Fisher was able to introduce randomization as a principle, which he integrated into statistical theory (Yates 1964).

It is important to take note that the early advances in the use of randomized controlled trials happened in the study of public policy as much as in psychology or the study of behaviour, suggesting a natural synergy between policy questions and randomized evaluations. As early as the 1920s, policy makers started to use experiments as knowledge about them increasingly disseminated. There were studies on the impact of milk on children's health (Pollock 2006). Many of these studies were defective because they did not clearly report how the children were allocated to the treatment. The first example was in Baltimore. Elmer McCollum, a nutritionist at the Johns Hopkins School of Hygiene and Public Health, carried out a field trial examining the health effects of supplementary milk on 84 children who were allocated to two groups of equal size; but the methods of allocation were not clearly reported and the diets changed during the research. Other early pilot work in Scottish schools led to the commissioning of the more ambitious Lanarkshire Schools Milk Experiment, which was carried out in the spring of 1930, but which was messed up by the teachers by varying the assignment (Student 1931). The failings of such experiments partly reflect the lack of dissemination of the principles of experimentation, which was probably because Fisher's work was not at first well known – his book on the design of experiments only came out in 1935. They also reflect the difficulties of carrying out experiments and ensuring that partners follow a protocol. Nonetheless, the 1920s were when experiments were discussed and randomization was more commonly thought about. Better-implemented and randomized experiments appeared several decades later, such as the Cambridge-Somerville intervention, carried out in 1939, which tested support for young offenders (McCord 1978).

The medical trial has been incredibly successful in the post-1945 period, with many tens of thousands being completed each year, becoming the official standard for almost all medical treatments and procedures, and in effect shaping understandings and uses of the trial. There have been over 350,000 trials done up to 2002 according to one estimate (Michalopoulos 2005: 11), and medicine and medical practice remains at the heart of systematic reviewing as in the Cochrane reviews.

Gradually the use of experiments has expanded out from medicine. Influential was the work of Campbell on the statistical properties of experiments and quasi-experiments (Campbell and Stanley 1963), and how to use them in social settings (Campbell 1957). An important period was the expansion of welfare policies in the 1960s and the demands for stronger evaluations (Greenberg, Linksz, and Mandell 2003). An early, prominent example in the US was the use of experiments to test for the effect of negative income tax, which originated out of the Office of Economic Opportunity and were initiated in New Jersey. The idea was suggested by Heather Ross who wrote a proposal to the Office for Economic Opportunity. A negative income tax is a benefit to people of low income. It takes account of welfare support to ensure that people are not disincentivized by the tax system. When seeking work, people often lose benefits that are not compensated by low-income work. The negative income tax, sometimes called tax credits, ensures a smooth progression of income as people move from welfare to work. The project intended to assign 1,000 households to negative income tax. The project was highly complex and required a lot of choices, in particular over the allocation of tax credits in the sample. The experiment led to other negative income tax experiments in North Carolina and Iowa, and also in Seattle and Denver. In spite of all the difficulties, the negative income tax intervention has also been thought of as the first large-scale social experiment using randomization, which stimulated the more general use of trials as a means to evaluate public policies (Greenberg, Linksz, and Mandell 2003: 119).

In the 1970s and 1980s, there was an expansion of the range of social programmes evaluated with randomized allocation, in particular welfare-to-work policies (Riccio and Bloom 2002), job training (Bloom 1984; Bloom et al. 1997) and re-employment bonuses (Robins and Spiegelman 2001). Another early productive area was crime with evaluations being done to test hot spots policing (Sherman and Weisburd 1995) and peer mentoring (Petrosino, Turpin-Petrosino, and Buehler 2003).

Greenberg and associates (Greenberg, Linksz, and Mandell 2003; Greenberg and Shroder 2004) have charted the rise of social experiments, mainly in the US, and their compendium, which has gone through three editions (Greenberg and Shroder 2004), lists the social experiments and gives summaries of each one. They identify 240 experiments carried out up to 2002. They set out three periods of social experimentation: era I (1964–74) of lengthy costly experiments looking at whole programmes; era II in which more modest experiments test for more incremental changes to these programmes; then era III of a return to large evaluations but this time carried out by state governments. They show the gradual increase in the use of the experiment over these periods. Given the continuing interest in trials, Haskins and Margolis (2014: 16), in their review of the use of evidence-based policy during the Obama presidency, consider these numbers to be ‘seriously out of date’, though of course they have probably reduced during the Trump presidency.

Education is another area of expansion of RCTs. Congress tied randomized evaluation to the release of federal funds in the US Education Sciences Reform Act of 2002. Partly as a result, Baron finds 90 studies using trials carried out from 2002–13 (see John 2017: 69). And there has been an upswing in the trials commissioned since 2008, for example of the Reading Recovery programme, the Success for All programme for education of students in elementary schools, and the impact of the Teach for America programme (John 2017: 127).

The growing interest in trials caused scholars to call for more social experimentation as a way of life for government agencies (Greenberg, Linksz, and Mandell 2003), which reaches back to the earlier argument for reforms as experiments made by Campbell (1969). However, such periodic advances are also followed by the realization of difficulties of implementation with trials (Berk, Smyth, and Sherman 1988), which dampens enthusiasm. The use of trials seems to go in cycles, though in aggregate with increasing use over time.

One additional area of expansion with policy implications has been in the development field, in particular in the evaluation of aid programmes, with centres of activity at research organizations including Abdul Latif Jameel Poverty Action Lab and Innovations for Poverty Action. These experiments have been large scale, such as Olken’s corruption experiment in Indonesia (Olken 2007). Trials have now become widely established as a method of evaluation by economists in this field and link to academic work in political economy (Humphreys and Weinstein 2009). These studies have involved close collaborations with policy makers, such as donor governments or aid agencies that deliver the interventions. They are often targeted at understanding local decision making, especially at local and village level, and thus offer understanding of the policy process to understand how implementation works, just like public policy scholars do with case study methods.

Mainstream political science had to wait until the late 1990s for experiments to be fully rediscovered (Druckman et al. 2006). There had always been some interest in experiments with those from a background in psychology and Morton and Williams (2010: 6–7) identify an interest in political science as evidenced by reviews and handbooks, as well as studies before the 2000s, though these were not often cited. And there were even Get-Out-the-Vote

experiments done in the 1920s (Gosnell 1926). The current interest in field experiments was stimulated by the landmark study conducted at Yale University by Gerber and Green (2000), which launched a large number of experimental studies of voter turnout, many carried out by students at Yale as well as by the authors (Green and Gerber 2015), encouraging the diffusion of trials into other areas in political science. Topics include radio broadcasts, campaign spending, positive and negative campaign messages, television campaign advertisements, and internet advertisements. There has been a growing interest in laboratory experiments that has complemented the interest in the field. Then there has been the use of the internet to generate large-N trials, such as through election mobilization, or seeking to understand mobilization through social media, such the mobilization study that went to 61 million users in the 2010 Congressional elections (Bond et al. 2012).

The US has led the way in the use of trials in public policy, but other countries have experimented with them too. One early example is the 1966 experiment in the UK on electricity pricing carried out between 1966 and 1972 (Greenberg and Shroder 2004: 439–41). This experiment involved testing different pricing schemes on 3,420 residents who got a seasonal tariff, 840 who received a seasonal time of day tariff, 840 who got a target amount and a control group of 900 that received block rates. Overall, the UK has had a patchy experience in the use of trials with the exception of the central government employment department (in its various incarnations), which has had a history of using trials going back to the 1980s, mainly testing initiatives to encourage employment. There was a gradual increase in the use of trials in the 1990s associated with the Labour government's evidence-based policy initiative, and the publication of the official handbook of evaluation, *The Magenta Book* (HM Treasury 2011), which highlighted trials as a superior form of evaluation. Other examples include the evaluation of the HM Prison Service Enhanced Thinking Skills Programme. One area of expansion has been in education, for example the Welsh government's free breakfast initiative, and then the work of the Education Endowment Foundation in funding trials. Health and social care is another example coordinated by the UK Department of Health.

It is only since 2010 that the interest in trials has really taken off, with most testing behaviour change initiatives. This research has been pioneered by the work of the Behavioural Insights Team (BIT), which has sought to show evidence for behavioural interventions. The team has carried out a number of experiments on reminders to pay taxes, court fines, energy savings, health and charitable giving, and has produced a guide about how to do experiments (Haynes, Goldacre, and Torgerson 2012). The team has prompted a wave of experimentation across and beyond government, and to the rest of the world (John 2018). At the same time there has been a spate of popular science books that have sought to extend the revolution in evidence-based medicine to public policy more generally, which includes an advocacy of trials (e.g. Henderson 2012). In spite of their concentration in the medical field, the patchy progress and the history of some failures of implementation, it is fair to say that randomized evaluations are much more commonly used in the applied social sciences and now by policy makers. The question becomes whether this practical interest turns into studies of the policy process itself.

5. ELITE EXPERIMENTS

There has been an expansion of the use of experiments in public policy, but mainly directed to evaluation and influencing individual behaviour. More recent work moved experiments closer

to the concerns of scholars interested in decision making within institutions. As much of political science is concerned with the operation of institutions and the representation of interests within the political system it is a natural site for experiments, but political institutions are usually hard to manipulate and are often in insufficient numbers to sustain statistical modelling. However, actors within institutions, the elected representatives, do exercise an important role in structuring those institutions to make policy choices. They receive and act on information from their environment so they can be manipulated. Their numbers are not as great as those done on citizens, but there are many elected representatives and even more bureaucrats to study. By examining legislator responses to constituent demands, it is possible to contribute to debates about equality and representation (Butler 2014) from how representatives respond to interest groups and campaigns or are held accountable when carrying out institutionally defined roles. It is possible to get insights into decision making and on how leaders balance interests. With bureaucrats, of key interest in the study of public policy, it is possible to test for their responsiveness. Even if the institution itself is not varied, experiments can help political scientists understand how the institution works in a different way from much observational research. As Grose argues, there is 'vast potential for field experiments in the study of political institutions' (Grose 2014: 356). One question that emerges is whether issues of sample size and manipulability limit the range of these experiments. The bigger question is whether the switch from relatively powerless citizens – who might write a complaining email or make a telephone call – to more powerful politicians and bureaucrats has implications for the range and type of experiments that can be done. It is also not possible to vary important signals that politicians and bureaucrats respond to, such as external crises, media scandals, strategies from opposing political parties, Supreme Court decisions, competition with other jurisdictions, and the like. On the other hand, experimenters have been ingenious in discovering new ways to test for elite responses, which the following sections of text will show.

It is also conventional to include elites in survey experiments. This procedure is about the assessment of vignettes or responses to hypothetical situations (e.g. Avellaneda 2013; Nielsen and Baekgaard 2015), which is different from field experiments where elites are responding to an actual rather than hypothetical intervention. It is advantageous to use elite members rather than students to deliver these kinds of survey experiments, as one can assume that the respondents are using insights from their actual roles and experience. It is possible to include a wide range of scenarios that correspond to the real situations that politicians find themselves in. But in the end, these are not typical field experiments and do not have the key advantage that the method offers, which is to simulate real conditions and observe actual responses by politicians or bureaucrats. The survey experiment offers a view from the politicians or bureaucrats about what they might do in response to a hypothetical scenario and in the full knowledge by the participants that they are part of a research project.

One further way of carrying out elite experiments is to target the public authority itself directly, through a communication that might be answered and handled by a bureaucrat or politician or both. Here, the researcher needs many authorities, which can usually only appear at the sub-national level or below that at a village or small community. Numerosity is one of the reasons why experiments studying the decision making of village elites have been used in developing contexts. Beath, Fotini, and Enikolopov (2013) randomized 50 villages in Afghanistan where one group required women's participation in elite decisions as a condition of the aid being delivered. The researcher monitored the outcomes. At first, this intervention might look like a policy experiment, and it is in part linked the performance of aid. But, it is

an elite experiment because the existing elites have to respond to the women who are new decision makers and can exercise power. Such experiments can also be done in the developed context. Worthy, John, and Vannoni (2017) sent a Freedom of Information request to 5,000 parish councils in England to test responsiveness, work that has been replicated in the Netherlands (Grimmelikhuisen et al. 2019). Ryan et al. (2018) randomly allocated an intervention to increase participation among English parishes. Butler (2010) randomly allocated 60 of Kentucky's counties to a control group and 60 to the treatment condition of receiving a letter from an interest group saying they were going to monitor their procedures for voter registration of young people. Hess, Hanmer, and Nickerson (2016) sent emails from a US state seeking to improve counties' implementation of the Voter Registration Act.

Field experiments on politicians are a relatively recent phenomenon. An early study is Chin, Bond, and Geva (2000), which recruited politicians to take part in an exercise designed to stimulate a response to a constituent. It is impressive because it is one of the few experiments involving Congressional politicians (through their staffers), but it is not a field experiment. Nonetheless, there is a rich vein of recruiting politicians to work with researchers, such as Sheffer et al.'s (2018) experiments to find out if politicians are better decision makers than citizens: they are not and have the same biases as citizens, which has large implications for the study of public policy given the long debate in studies of the policy process, about whether decision makers conform to the rational model or whether a more behavioural model is a better account, which goes back to Simon (1955) and extends from Allison (1971) to Baumgartner and Jones (2010). There is now a rich vein of work emerging, such as an experiment in the Netherlands to test whether politicians behave according to prospect theory in taking risks, which they do, but slightly less than the general population (Linde and Vis 2017).

The main example of an elite field experiment is mass emails or letters. Butler (2014: 23) attributes the idea to Putnam (Putnam, Leonardi, and Nanetti 1994) whose research team sent messages to Italian regional bureaucrats to measure their responsiveness. This activity was not an experimental intervention, however. Probably the first such is Bergan (2009), who worked with a campaign organization sending emails to members of the House of Representatives in New Hampshire. They used the web software of the activists to send messages directed to the randomly allocated elected representatives of a treatment group, which contrasted with the control group not receiving any. They showed that the emails influence legislative behaviour. Such is the size of this research area, a meta-analysis of studies of the bias of officials has recently been published (Costa 2017).

An important vein of work examines the responsiveness of legislators to being watched or provided information by an external source. Butler and Nickerson (2011) sent randomly selected legislators the poll results of their constituents' policy preferences, showing that those in the treatment groups were more likely to vote in line with their voters' preferences. Being watched has an effect as shown by Nyhan and Reifler (2015) who fact-checked legislators in nine US states.

The most prominent example is the experiment done by Butler and Broockman (2011) who sent emails from fake people to assess the responsiveness of state legislators. They varied the implied race of the respondent's name to show bias in responsiveness. There are similar studies, such as McClendon's (2016) research on local councillors in South Africa. Butler (2014) has also carried out responsiveness experiments on mayors and tests for responsiveness according to gender and social economic status of the senders.

Research on bureaucrats has been less common than that on politicians, which is a surprise as there are large numbers of them and experiments on them are less likely to suffer from contamination if they are spread across many agencies (e.g. local governments). Now the bureaucracy is a growing area of experimental research. One recent example (White, Nathan, and Faller 2015) examined the responsiveness of election officials to a request to increase voter registration among young people; another study found those election officials sensitive to the partisan sources of the contact (Porter and Rogowski 2018). Einstein and Glick (2017) sent 1,000 fake emails asking about welfare provision using names from different racial groups, though this example is less an experiment on elites than frontline bureaucrats. A recent set of experiments seeks to throw light on the implementation process, which again is an important part of the policy process and of interest to policy scholars. For example, experiments in four countries show whether the policy positions of bureaucrats at the frontlines of government are susceptible to frames and cues from communications (Andersen and Jakobsen 2017). There are relatively few lobbying experiments with politicians, but Richardson and John (2012) worked with local lobby groups to send differently worded letters to randomly allocated local politicians in England.

The use of trials in the academic field of public administration started modestly, and trials were first carried out in the laboratory, as survey experiments or in simulations or vignettes presented to policy makers, which may be the result of the difficulty of agreeing and doing them, such as James and Moseley's (2014) study of the use of performance information, and a set of Danish trials examining performance management and policy (e.g. Andersen and Moynihan 2016). More recent experiments are carried out in collaboration with policy makers. Overall, there is a growing interest in public administration in all kinds of trials, including those done in the field (James and Moseley 2014), which has lessons for students of public policy about what can be achieved (see James, Gilke, and Van Ryzin 2017).

6. CONCLUSIONS: LESSONS FOR STUDENTS OF THE POLICY PROCESS

Experiments have been useful for policy evaluation, but until recently scholars of the policy process could be excused for thinking that they do not have much relevance for studies of the policy process, mainly because the complexity of public policy does not offer neat tests, such as whether an intervention works or not, or whether citizens might react to endorsement, for example. Yet the recent development of elite experiments in political science shows how experiments can be used to understand decision making more generally. If experiments do not offer holistic understandings of the policy process, nor reveal how a policy is made, they can test aspects of decision making that can cumulate over time into a set of findings that would offer more insights into decision making. That the allied subject of public management should move so fast to understanding decision making by bureaucrats is a sign that there are ample things to test. There are some signs that quasi-experimental methods are making an appearance in public policy, such as the use of difference-in-differences as a method to test for feedback in tobacco policy (Vannoni 2019). Another development is the use of natural and experimental cut-offs to examine the impact of changes to policy design, especially as most citizen/policy maker interfaces are digital. Hale et al.'s (2018) study of the impact of platform changes on citizen petitions, argues that changes can lead to more volatility, thus contributing to a more

unstable policy environment. The key idea is that petition signing before and after the date can be analysed as if they were randomly allocated so generating the causal inference. More of course can and should be done, carrying out randomized trials, leveraging natural experiments, and using quasi-experimental methods.

NOTE

1. A more in-depth treatment of the design and history of experiments is found in John (2017).

REFERENCES

- Allison, G. T. (1971). *Essence of Decision: Explaining the Cuban Missile Crisis*. Boston, MA: Little, Brown.
- Andersen, S. C. and Jakobsen, M. (2017). Policy positions of bureaucrats at the front lines: are they susceptible to strategic communication? *Public Administration Review* 77 (1): 57–66.
- Andersen, S. C. and Moynihan, D. P. (2016). Bureaucratic investments in expertise: evidence from a randomized controlled field trial. *Journal of Politics* 78 (4): 1032–44.
- Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Avellaneda, C. N. (2013). Mayoral decision-making: issue salience, decision context, and choice constraint? An experimental study with 120 Latin American mayors. *Journal of Public Administration Research and Theory* 23 (3): 631–61.
- Baumgartner, F. R. and Jones, B. D. (2010). *Agendas and Instability in American Politics*. Chicago, IL: University of Chicago Press.
- Beath, A., Fotini, C., and Enikolopov, R. (2013). Empowering women through development aid: evidence from a field experiment in Afghanistan. *American Political Science Review* 107 (3): 540–557.
- Berk, R. A., Smyth, G. K., and Sherman, L. W. (1988). When random assignment fails: some lessons from the Minneapolis spouse abuse experiment. *Journal of Quantitative Criminology* 4 (3): 209–23.
- Bloom, H. S. (1984). Estimating the effect of job-training programs, using longitudinal data: Ashenfelter's findings reconsidered. *Journal of Human Resources* 19 (4): 544–56.
- Bloom, H. S., Orr, L. L., Bell, S. H., Cave, G., Doolittle, F., Lin, W., and Bos, J. M. (1997). The benefits and costs of JTPA Title II-A programs: key findings from the National Job Training Partnership Act Study. *Journal of Human Resources* 32 (3): 549–76.
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., and Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature* 489 (7415): 295–8.
- Butler, D. M. (2010). Monitoring bureaucratic compliance: using field experiments to improve governance. *Public Sector Digest* (Winter): 41–4.
- Butler, D. M. (2014). *Representing the Advantaged: How Politicians Reinforce Inequality*. Cambridge: Cambridge University Press.
- Butler, D. M. and Broockman, D. E. (2011). Do politicians racially discriminate against constituents? A field experiment on state legislators. *American Journal of Political Science* 55 (3): 463–77.
- Butler, D. M. and Nickerson, D. W. (2011). Can learning constituency opinion affect how legislators vote? Results from a field experiment. *Quarterly Journal of Political Science* 6 (1): 55–83.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* 54 (4): 297–312.
- Campbell, D. T. (1969). Reforms as experiments. *American Psychologist* 24 (4): 409–29.
- Campbell, D. T. and Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Cartwright, N. and Hardie, J. (2012). *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press.

- Chin, M. L., Bond, J. R., and Geva, N. (2000). A foot in the door: an experimental study of PAC and constituency effects on access. *Journal of Politics* 62 (2): 534–49.
- Chong, A., De La O, A. L., Karlan, D., and Wantchekon, L. (2015). Does corruption information inspire the fight or quash the hope? A field experiment in Mexico on voter turnout, choice, and party identification. *Journal of Politics* 77 (1): 55–71.
- Deaton, A. and Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine* 210: 2–21.
- Druckman, J. N., Green, D. P., Kuklinski, J. H., and Lupia, A. (2006). The growth and development of experimental research in political science. *American Political Science Review* 100 (4): 627–35.
- Dunning, T. (2012). *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge: Cambridge University Press.
- Einstein, K. L. and Glick, D. M. (2017). Does race affect access to government services? An experiment exploring street-level bureaucrats and access to public housing. *American Journal of Political Science* 61 (1): 100–116.
- Fearon, J. D., Humphreys, M., and Weinstein, J. M. (2009). Can development aid contribute to social cohesion after civil war? Evidence from a field experiment in post-conflict Liberia. *American Economic Review* 99 (2): 287–91.
- Fisher, R. A. (1926). The arrangement of field experiments'. In S. Kotz and N. L. Johnson (Eds.), *Breakthroughs in Statistics: Methodology and Distribution*. New York: Springer, 82–91.
- Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.
- Gerber, A. S. and Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: a field experiment. *American Political Science Review* 94: 653–63.
- Gerber, A. S. and Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.
- Glennester, R. and Takavarasha, K. (2013). *Running Randomized Evaluations: A Practical Guide*. Princeton, NJ: Princeton University Press.
- Gosnell, H. F. (1926). An experiment in the stimulation of voting. *American Political Science Review* 20 (4): 869–74.
- Green, D. P. and Gerber, A. S. (2015). *Get Out the Vote: How to Increase Voter Turnout*. Washington, DC: Brookings Institution Press.
- Greenberg, D. H., Links, D., and Mandell, M. (2003). *Social Experimentation and Public Policymaking*. Washington, DC: The Urban Institute Press.
- Greenberg, D. H. and Shroder, M. (2004). *The Digest of Social Experiments*. Washington, DC: The Urban Institute Press.
- Grimmelikhuijsen, S., John, P., Meijer, A., and Worthy, B. (2019). Do freedom of information laws increase transparency of government? A replication of a field experiment. *Journal of Behavioral Public Administration* 2 (1). <https://doi.org/10.30636/jbpa.12.34>.
- Grose, C. R. (2014). Field experimental work on political institutions. *Annual Review of Political Science* 17: 355–70.
- Hale, S. A., John, P., Margetts, H., and Yasserli, T. (2018). How digital design shapes political participation: a natural experiment with social information. *PLoS ONE* 13 (4): e0196068.
- Haskins, R. and Margolis, G. (2014). *Show Me the Evidence: Obama's Fight for Rigor and Results in Social Policy*. Washington, DC: Brookings Institution Press.
- Haynes, L., Goldacre, B., and Torgerson, D. (2012). *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*. London: Cabinet Office.
- Henderson, M. (2012). *The Geek Manifesto: Why Science Matters*. New York: Random House.
- Hess, D. R., Hanmer, M. J., and Nickerson, D. W. (2016). Encouraging local compliance with federal civil rights laws: field experiments with the National Voter Registration Act. *Public Administration Review* 76 (1): 165–74.
- HM Treasury (2011). *The Magenta Book: Guidance for Evaluation*. London: HM Treasury.
- Humphreys, M. and Weinstein, J. W. (2009). Field experiments and the political economy of development. *Annual Review of Political Science* 12: 367–78.
- James, O., Jilke, S., and Van Ryzin, G. (Eds.) (2017). *Experiments in Public Administration Research: Challenges and Contributions*. Cambridge: Cambridge University Press.

- James, O. and Moseley, A. (2014). Does performance information about public services affect citizens' perceptions, satisfaction, and voice behaviour? Field experiments with absolute and relative performance information. *Public Administration* 92 (2): 493–511.
- John, P. (2017). *Field Experiments in Political Science and Public Policy: Practical Lessons in Design and Delivery*. New York: Routledge.
- John, P. (2018). *How Far to Nudge? Assessing Behavioural Public Policy*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Jones, B. D. and Baumgartner, F. R. (2005). *The Politics of Attention: How Government Prioritizes Problems*. Chicago, IL: University of Chicago Press.
- Linde, J. and Vis, B. (2017). Do politicians take risks like the rest of us? An experimental test of prospect theory under MPs. *Political Psychology* 38 (1): 101–17.
- McClendon, G. H. (2016). Race and responsiveness: an experiment with South African politicians. *Journal of Experimental Political Science* 3 (1): 60–74.
- McCord, J. (1978). A thirty-year follow-up of treatment effects. *American Psychologist* 33 (3): 284–9.
- Michalopoulos, C. (2005). Precedents and prospects for randomized experiments. In H. S. Bloom (Ed.), *Learning More from Social Experiments: Evolving Analytic Approaches*. New York: Russell Sage Foundation, 1–36.
- Morton, R. B. and Williams, K. C. (2010). *Experimental Political Science and the Study of Causality: From Nature to the Lab*. New York: Cambridge University Press.
- Nielsen, P. A. and Baekgaard, M. (2015). Performance information, blame avoidance, and politicians' attitudes to spending and reform: evidence from an experiment. *Journal of Public Administration Research and Theory* 25 (2): 545–69.
- Nyhan, B. and Reifler, J. (2015). The effect of fact-checking on elites: a field experiment on US state legislators. *American Journal of Political Science* 59 (3): 628–40.
- Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy* 115 (2): 200–249.
- Petrosino, A., Turpin-Petrosino, C., and Buehler, J. (2003). Scared straight and other juvenile awareness programs for preventing juvenile delinquency: a systematic review of the randomized experimental evidence. *Annals of the American Academy of Political and Social Science* 589 (1): 41–62.
- Pollock, J. (2006). Two controlled trials of supplementary feeding of British school children in the 1920s. *Journal of the Royal Society of Medicine* 99 (6): 323–7.
- Porter, E. and Rogowski, J. C. (2018). Partisanship, bureaucratic responsiveness, and election administration: evidence from a field experiment. *Journal of Public Administration Research and Theory* 28 (4): 602–17.
- Riccio, J. A. and Bloom, H. S. (2002). Extending the reach of randomized social experiments: new directions in evaluations of American welfare-to-work and employment initiatives. *Journal of the Royal Statistical Society Series A: Statistics in Society* 165 (1): 13–30.
- Richardson, L. and John, P. (2012). Who listens to the grass roots? A field experiment on informational lobbying in the UK. *British Journal of Politics and International Relations* 14 (4): 595–612.
- Robins, P. K. and Spiegelman, R. G. (2001). *Reemployment Bonuses in the Unemployment Insurance System: Evidence from Three Field Experiments*. Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.
- Ryan, M., Stoker, G., John, P., Moseley, A., James, O., Richardson, L., and Vannoni, M. (2018). How best to open up local democracy? A randomised experiment to encourage contested elections and greater representativeness in English parish councils. *Local Government Studies* 44 (6): 766–87.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Shadish, W. R., Cook, T. D., and Leviton, L. C. (1991). *Foundations of Program Evaluation: Theories of Practice*. London: Sage.
- Sheffer, L., Loewen, P. J., Soroka, S., Walgrave, S., and Sheaffer, T. (2018). Nonrepresentative representatives: an experimental study of the decision making of elected politicians. *American Political Science Review* 112 (2): 302–21.
- Sherman, L. W. and Weisburd, D. (1995). General deterrent effects of police patrol in crime 'hot spots': a randomized, controlled trial. *Justice Quarterly* 12 (4): 625–48.

- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics* 69 (1): 99–118.
- Student (1931). The Lanarkshire milk experiment. *Biometrika* 23 (3/4): 398–406.
- Torgerson, D. J. and Torgerson, C. J. (2008). *Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction*. Basingstoke: Palgrave Macmillan.
- Vannoni, M. (2019). A behavioral theory of policy feedback in tobacco control: evidence from a difference-in-difference-in-difference study. *Policy Studies Journal* 47 (2): 353–71.
- White, A. R., Nathan, N. L., and Faller, J. K. (2015). What do I need to vote? Bureaucratic discretion and discrimination by local election officials. *American Political Science Review* 109 (1): 129–42.
- Worthy, J., John, P., and Vannoni, M. (2017). ‘Transparency at the parish pump: a field experiment to measure the effectiveness of freedom of information requests in England’. *Journal of Public Administration Research and Theory* 27 (3): 485–500.
- Yates, F. (1964). Sir Ronald Fisher and the design of experiments. *Biometrics* 20 (2): 307–21.

10. Measuring change in comparative policy analysis: concepts and empirical approaches

Jale Tosun and Julia Schnepf

1. INTRODUCTION

The motivations for reforming public policies have been a major research interest of comparative policy analysis for decades, starting with the work by Hogwood and Peters (1982), but then culminating in a fully fledged research agenda (e.g. Peters, Fontaine, and Mendez 2018). The influential work of Hall (1993) has prepared the ground for numerous insightful empirical studies of policy change, which adopt both a qualitative and a quantitative approach (e.g. Fontaine, Narvaez, and Velasco 2018; Holzinger, Knill, and Arts 2008; Knill, Schulze, and Tosun 2012; Knill, Tosun, and Heichel 2008; Mendez 2018; Pralle 2006; Tosun 2013). Building on Hall, several studies have discussed this phenomenon conceptually, asking about the different manifestations of policy change (e.g. Capano 2009; Howlett and Cashore 2009) such as policy expansion, policy dismantling, and policy termination (e.g. Geva-May 2004; Jensen et al. 2014; Jordan, Bauer, and Green-Pedersen 2013; Knill, Schulze, and Tosun 2012). Further to this, policy change has been central to policy process theories (Weible and Sabatier 2017), including the Advocacy Coalition Framework (e.g. Nohrstedt and Weible 2010), the Punctuated Equilibrium Theory (Baumgartner and Jones 2009 [1993]), the Multiple Streams Approach (Kingdon 2002; see also Cairney and Jones 2016), and the Policy Subsystem Adjustment Model (Howlett and Ramesh 2002). In addition to these perspectives, there exist many more focused theories that concentrate, for example, on policy learning (e.g. Dunlop and Radaelli 2013), the role of lobbying (e.g. Baumgartner et al. 2009), or how political leadership matters (e.g. Mendez 2018).

The operationalization of policy change warrants enhanced attention since how we measure policy change will affect what conclusions we draw from the respective analysis concerning the validity of our theoretical reasoning. While there is general agreement that the measurement of policy change requires at least the measurement of policies at two points in time, characterizing policy change across countries or sub-national units is a challenging endeavor. In fact, one of the key challenges of measuring policy change is the need to gather and analyze longitudinal data, which can be difficult to obtain, especially for a larger set of units. Another complicating factor is the fact that “policies increasingly come in complex packages” (Howlett and del Rio 2015: 1233), which makes the operationalization of policy change even more demanding and raises a new issue for methodological debate.

Despite these challenges, comparative measurement is necessary to advance the empirical testing of the theories of policy change (see Tosun and Workman 2017), and the prolific literature in comparative policy studies is a useful starting point for this endeavor. Building on that literature, in this chapter, we introduce and discuss ways of measuring policy change. We pursue the overall aim of demonstrating the richness of the literature, while at the same time

drawing attention to the consequences of the different measurement approaches for the study of policy change.

For the sake of a coherent presentation, we limit our treatise to policies in the field of environmental protection. In this way, we have variation in the data presentation, but we also ensure that all policies relate to the same policy domain, which facilitates the comparative assessment of the different measurements. In addition, with regard to the geographical scope of our analysis, we concentrate on the Americas, that is, North, Central, and South America. Our empirical focus is informed by several considerations. The United States was – and in some areas still is – a first-mover in environmental policy (Kochtcheeva 2009; Vogel 1986). Therefore, we can obtain longitudinal data for US environmental policy for a relatively long observation period. We assume that similar to the United States, Canada introduced environmental policies at a rather early point in time, making it a suitable case for observing policy change (e.g. Pralle 2006). The Central and South American countries have experienced a transition from closed economies to open and globally integrated market economies, which entailed increased economic activity and therefore more stress on the environment and the need to change existing policy arrangements or to introduce new policies (Tosun 2013). Therefore, we are confident to observe policy change data for these countries. Central and South America are also two regions affected by the third wave of democratization (Hagopian and Mainwaring 2005), which we also expect to have changed the need for policy makers to respond to public demands for environmental protection (Tosun 2013).

In this chapter, we proceed as follows. First, we give an overview of the concept of policy change to set the stage for the subsequent empirical analyses. Then, we present the three main approaches to the measurement of policy change and subsequently discuss their respective strengths and weaknesses. Subsequently, to complement our presentation, we also briefly outline the outcome approach to measuring policy change. In the final section, we summarize the main insights and offer some concluding remarks. Overall, we show that comparative public policy has made good progress toward the measurement of policy change. Nonetheless, we invite future research to invest more in aligning the conceptual discussions with empirical work.

2. CONCEPTUALIZING POLICY CHANGE

There have been numerous attempts to conceptualize policy change in policy studies, with the special issue of the *Journal of Comparative Policy Analysis* edited by Peters et al. (2018) representing one of the most recent endeavors. While the newer conceptual and methodological literature has provided many intriguing insights, any empirical approach to the study of policy change must position itself with regard to the influential conceptual framework put forward by Hall (1993). While Hall's work has predominantly guided the empirical analysis of policy change (e.g. Holzinger, Knill, and Arts 2008; Knill, Tosun, and Heichel 2008; Knill, Schulze, and Tosun 2012; Tosun 2013), the author actually puts forward a theoretical argument about policy learning – one of the dominant explanations for the occurrence of policy change (see also Weible and Sabatier 2017).

Hall's approach differentiates between three possible types of policy change: First-order change is about modifications to the precise *setting* of a policy instrument. Second-order change refers to the variation in policy *instruments*. Third-order change is about changes in the overarching *paradigm* of a policy (e.g. Fontaine, Narvaez, and Velasco 2018). Concerning the

latter, Holzinger and Knill (2008) suggested replacing this dimension with the more pragmatic concept of policy *presence*, that is, whether a country has developed a specific policy or not with regard to a certain problem. With regard to the three dimensions of Hall's concept, Knill, Schulze, and Tosun (2012) advocate the inclusion of a fourth dimension that captures the "scope" of a policy, that is, which individuals, companies, products, processes, or territorial units are covered by a given policy. Through this extension, the authors argue, the assessment of policy change becomes more fine-grained and allows for capturing more instances of change. Another merit of this refined approach is that the study of policy change can be linked with the literature on target population as suggested by Schneider and Ingram (1993), which provides a novel theoretical perspective on the causes of policy change.

Returning to the original framework by Hall (1993), policy change is most likely to occur for the dimension of the policy setting (i.e. first-order change), but it becomes more difficult with regard to policy instruments (i.e. second-order change) and even more so with policy paradigms (i.e. third-order change). Changes in settings are considered as rather easy to attain since they can be achieved within existing instruments and paradigms. Similarly, instrumental changes are considered to be more likely than paradigm changes as the latter imply the departure from dominant and institutionally strongly entrenched ideas of how to perceive and resolve certain societal problems (e.g. Béland 2009; Fontaine, Narvaez, and Velasco 2018).

From this perspective, we consider third-order change as "major" policy change whereas second-order and first-order change correspond to "moderate" and "minor" policy change, respectively.

A different conceptual approach is put forth by Baumgartner and Jones (2009 [1993]), who are also interested in studying policy change, but by paying more attention to the politics dimension (Tosun and Workman 2017). Rather than examining in detail the quality of policy change, this approach is interested in gauging broader patterns concerning quality and quantity. The Punctuated Equilibrium Theory, which Baumgartner and Jones introduced to political science, contends that although generally marked by stability and incrementalism (i.e. minor policy change), political processes occasionally produce large-scale departures from the past (i.e. policy punctuations; see also John and Bevan 2012). Thus, this approach is predominantly interested in describing patterns of policy change, which entails the need for large quantities of data for a long period.

Similar to Hall's approach, the Punctuated Equilibrium Theory is not interested in empirical patterns only, but in explaining how they come about. To this end, Baumgartner and Jones conceive policy makers to be limited in their resources so that they cannot consider all problems and their (potential) solutions at all times. Therefore, policy makers tend to ignore problems, but sometimes they pay them great amount of attention (e.g. Maor, Tosun, and Jordan 2017; Peters, Jordan, and Tosun 2017). Which issues receive attention depends on (interest) groups and their efforts to place them on the agenda (or keep them off the agenda) by framing the issues in a specific way. Policy change becomes feasible when (interest) groups challenge the way in which policy makers understand a problem. For example, (interest) groups can shift an issue to a new policy venue that is composed of a different set of actors, who might be more receptive of that issue (Baumgartner and Jones 2009 [1993]).

The main interest of diffusion research lies in how policy innovations spread from one entity to another (Gray 1973; Walker 1969), which then leads to policy change in the respective entities adopting the innovations. While straightforward at first glance, this definition of a policy innovation requires further elaboration. The first need for clarification concerns the

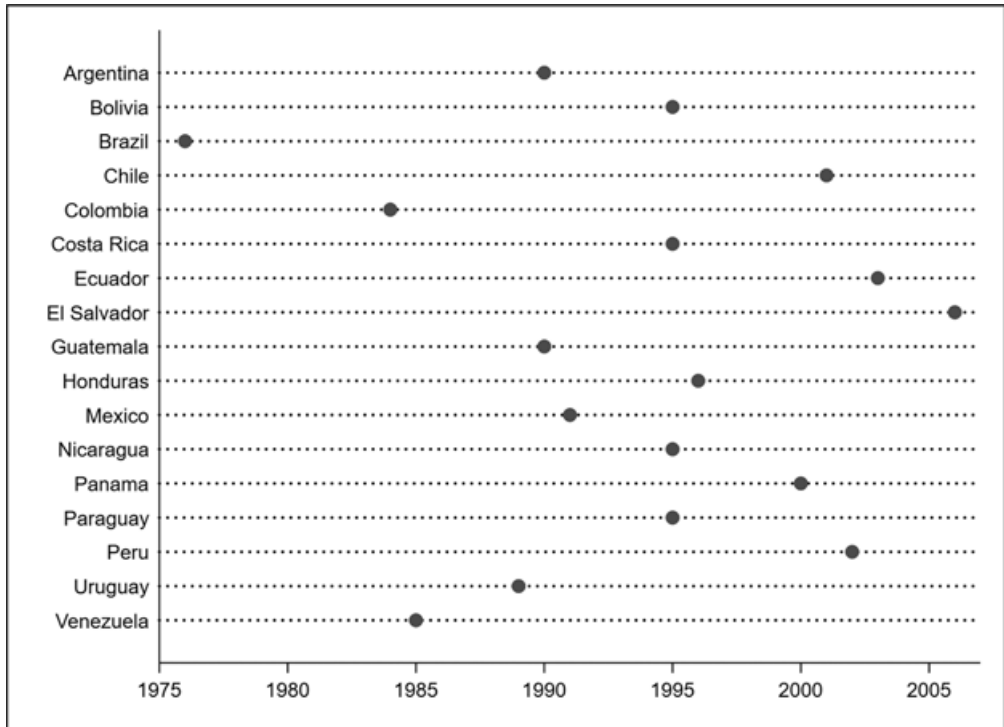
question about what characterizes a policy innovation. According to Walker (1969), a policy innovation is about a national government adopting a new policy, while “new” means that it is new to the jurisdiction in question (see also Tosun 2018). We can establish a link between diffusion research and the conceptual framework by Hall (1993) as modified by Holzinger and Knill (2008). From that perspective, policy diffusion corresponds to change with regard to the dimension of policy presence: the policy innovation did not exist in a jurisdiction and is introduced for the first time by means of policy adoption. Policy diffusion could also correspond to third-order change as defined by Hall (1993), but this requires the policy innovation concerned to introduce a new policy paradigm to a country.

Diffusion studies typically begin by offering a characterization of policy adoption patterns based on the cumulative number of countries or entities that have adopted a given policy by time t . In most cases, this produces an S-shaped curve, implying that adoption is slow at first, then rapid, and finally levels off as saturation is reached (Gray 1973). In a next step, diffusion scholars seek to provide an explanation for the observation of the typical S-shaped curve. In this regard, a consistent set of so-called diffusion mechanisms has emerged, which focuses on the role of emulation, learning, and competition (e.g. Biesenbender and Tosun 2014; Dobbin, Simmons, and Garrett 2007; Maggetti and Gilardi 2016).

In the next sections, the three perspectives presented above will guide our overview of the measurement of policy change. We acknowledge that these perspectives are a selection from a greater set of concepts that have emerged over the last decades. Nonetheless, we are confident that our selective approach is still instructive since these are the most established measurements in the literature on comparative policy analysis (see Tosun and Workman 2017). In addition, the concepts selected are complementary: While the measurement approach underlying the Punctuated Equilibrium Theory allows for describing patterns of policy change, Hall’s (1993) approach and its modifications by Holzinger and Knill (2008) and Knill, Schulze, and Tosun (2012) give us an opportunity for an in-depth examination of policies and how they change over time. With the diffusion perspective, we can concentrate on the first-time adoption of a policy innovation in a jurisdiction.

3. THE MULTI-DIMENSIONAL APPROACH TO MEASURING POLICY CHANGE

We begin our illustration of the empirical application of the concepts of policy change presented in the previous section with the approach by Hall (1993), which we refer to as the multi-dimensional approach. However, for the first empirical impression of this approach, we rely on the simplified definition of third-order change by Holzinger and Knill (2008). Therefore, Figure 10.1 gives an overview of the year in which 17 countries in Central and South America for the first time ever adopted a standard that defines limit values for the organic pollution of surface waters, which corresponds to the notion of policy presence. An emissions standard is a classic command and control instrument: The “command” aspect relates to the setting of the maximum level of permissible pollution, whereas the “control” aspect is about monitoring and enforcing the level defined by means of fines and other forms of sanctions. In conceptual terms, Figure 10.1 is instructive since it demonstrates that when applied empirically, the presence of a policy is also accompanied by the definition of a given policy instrument, which indicates that the dimensions identified by Hall (1993) and modified



Source: Own elaboration based on Tosun (2013).

Figure 10.1 First-time adoption of emission standards for organic water pollutants

by Holzinger and Knill (2008) and Knill, Schulze, and Tosun (2012) are not independent from each other. This might be a desirable feature, but it also entails that there exist interdependencies that need to be taken into account when analyzing the data.

In substantive terms, Figure 10.1 shows that there is a considerable gap between the country that first adopted an emission standard for organic water pollution (Brazil in 1976) and the last one (El Salvador in 2006) to adopt it. Based on this data, we can state that all countries experienced a change in the presence of this particular policy. What we cannot infer from the figure is how strictly the emission standards are defined and neither can we see whether the standards have been changed after their first-time adoption. Before we examine these aspects, we need to go back to the original definition of third-order change by Hall (1993).

Determining paradigmatic change is not an easy endeavor since according to Hall it is not only about the presence of a new policy, but there must also be a new idea underlying it (see Béland 2009). Therefore, it does not come as a surprise that many studies of paradigmatic change rely on a qualitative approach (e.g. Fontaine, Narvaez, and Velasco 2018). We follow the dominant approach adopted in the literature and attempt to illustrate the logic of a third-order change by an in-depth case description.

A suitable case for that purpose is the policy on green biotechnology in Brazil and how it changed. In 1995, the center-right coalition of President Fernando Cardoso adopted the first policy to regulate the release of genetically modified organisms (GMO) into the environment.

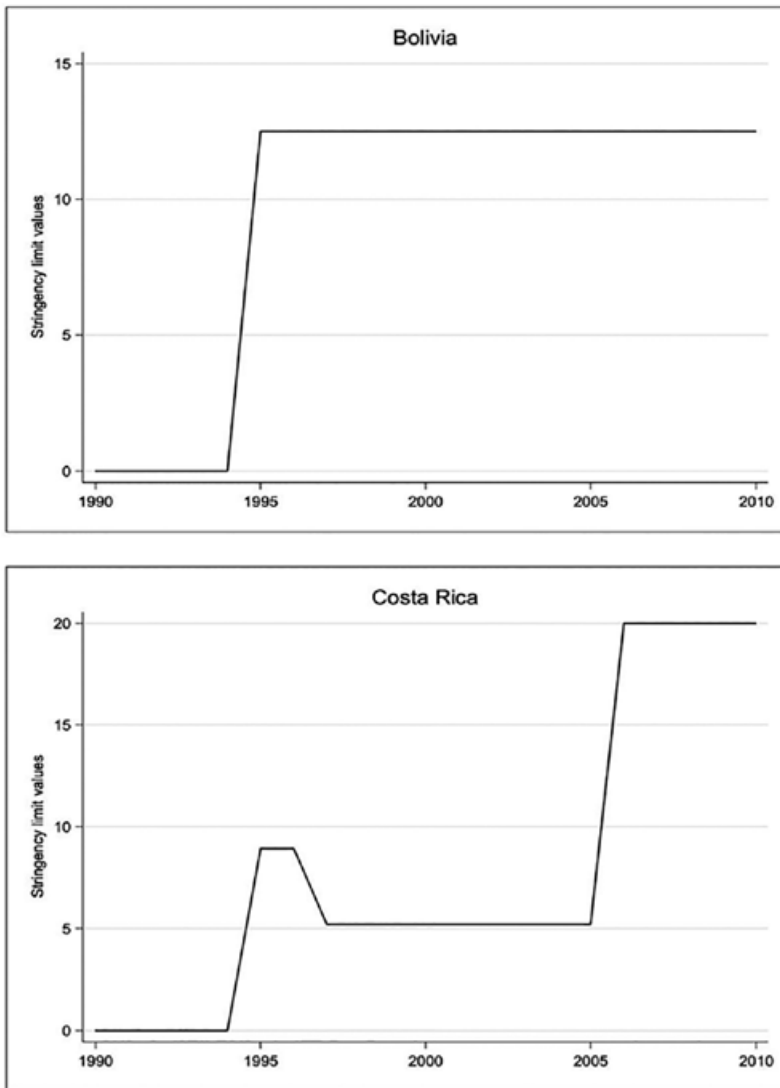
The Biosafety Law permitted GMO-related research as well as the sale of commercial products containing GMOs, but it prohibited the commercial cultivation of GM crops (Jepson, Brannstrom, and Stancato de Souza 2008). In 1998, the entry into the Brazilian market of the US-based company Monsanto challenged this policy paradigm. The company requested permission from the Brazilian government to commercialize its GM crop seeds with the goal of promoting the commercial cultivation of GM crops in the country. A paradigmatic policy change took place with the entering into office of President Luiz Inácio Lula da Silva in 2003. The Lula administration deviated from the previous policy and allowed the commercial cultivation of GM crops with the result that today Brazil is one of the major GM crop producers in the world (Tosun 2013). Thus, the idea underlying the Brazilian policy on green biotechnology became transformed, shifting from a policy regime that prohibited the commercial cultivation of GMOs to one where this is allowed.

We can also make use of an in-depth study to illustrate an instance of second-order change. For this purpose, mining policy in Peru offers a good example. In the last years, there have been intensive environment-related conflicts between mining project developers and the communities affected by their activities. When seeking authorization for a mining project, companies must present an Environmental Impact Assessment. However, in response to the intensification of the conflicts at the local level, the government adopted a new policy instrument that asks mining project developers to organize and participate in workshops and public hearings (OECD 2016).

We now return to our first example of emissions standards for the pollution of surface water with organic pollutants. Figure 10.2 provides insights into how the limit values for organic water pollution changed over time in two exemplary countries: Bolivia and Costa Rica. Both countries introduced limit values for organic water pollution in 1995, but this is where the similarities end. When Bolivia first adopted the limit values for organic water pollution, it adopted a stricter approach that allowed a lower concentration of effluents to be emitted into water (measured by 12.5 regulatory units; for details, see Tosun 2013). Costa Rica started with a more lenient regulatory setting of 8.9 regulatory units, which was modified in 1997 to become even more lax (5.2 regulatory units), only to be changed another time to become much stricter than the Bolivian standard. More precisely, the emissions standards that were in place in Costa Rica in 2010 corresponded to 20 regulatory units compared to 12.5 regulatory units in Bolivia.

According to Hall's reasoning, changes in the setting of a policy instrument (i.e. first-order change) should be easy to attain and therefore should occur more frequently than the other forms of policy change. Nonetheless, Bolivia represents a case where policy makers abstained from changing the limit values again after they had introduced them. Therefore, we can conclude that the occurrence of policy change depends on many factors, even in the case of first-order change.

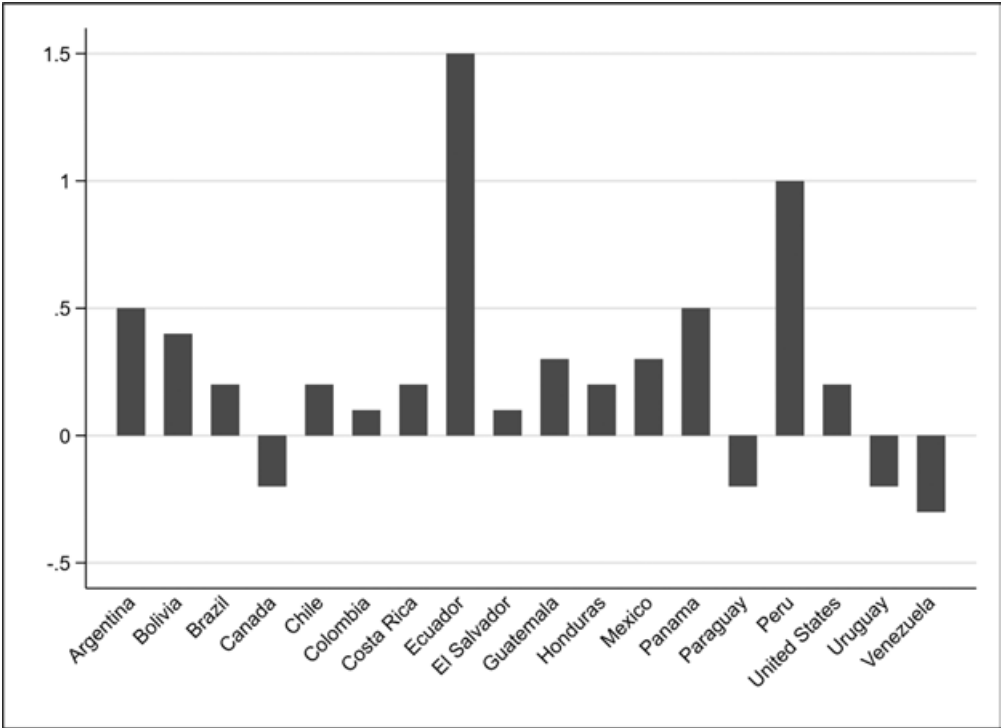
All examples given above relied on so-called policy output data, that is, legal acts actually adopted by policy makers (see Holzinger and Knill 2008; Holzinger, Knill, and Arts 2008; Knill, Tosun, and Heichel 2008; Knill, Schulze, and Tosun 2012; Tosun 2013). Analyzing that data has the advantage that we can adequately capture the factual decision making. Drawbacks of that data are that it is not readily available and that we need to adjust the data in order to make it comparable across countries and over time. While with regard to internal validity, this measurement is worth the effort, there also exist more pragmatic ways for assessing first-order change. For example, one can use survey data.



Source: Own elaboration based on Tosun (2013).

Figure 10.2 Changes in the stringency of limit values for organic water pollutants

The World Economic Forum's Executive Opinion Survey (Browne and Geiger 2009; Browne et al. 2016) provides survey data collected from top business and government leaders. The survey offers an item battery on perceptions of regulative policies for different environmental media. In addition to these specific items, the respondents are asked about the general stringency of environmental regulations in their countries, which allows for gauging perceived changes in the setting of multiple policy instruments as well as policy packages (see Howlett



Source: Own elaboration based on the executive opinion survey of the Global Competitiveness Reports published by the World Economic Forum.

Figure 10.3 *Changes in the perceived stringency of environmental policy between 2009 and 2016*

and del Rio 2015). The answers are coded using a Likert scale ranging from 1 (= very lax/ among the worst in the world) to 7 (= very rigorous/among the world’s most stringent).

Figure 10.3 presents data on the perceived general stringency of environmental regulations in 18 selected countries in North, Central, and South America. The data are taken from the 2009 (Browne and Geiger 2009) and the 2016 (Browne et al. 2016) releases of the Executive Opinion Survey. When inspecting the data, it becomes apparent that the average perception of regulatory stringency across countries has changed between these two points in time. The respondents indicate that they perceive all but four countries to have stricter environmental regulations in place in 2016 compared to 2009. Of the countries that were perceived to have improved their environmental regulations, Ecuador and Peru stick out since the difference between the scores for 2009 and 2016 is greatest. Somewhat surprisingly, among the four countries that were perceived to have less stringent environmental regulations in place in 2016 is Canada along with Paraguay, Uruguay, and Venezuela. There could be substantive reasons that explain the changes in the perceptions (e.g. adoption of stricter regulation) as well as methodological reasons such as changes in the composition of the groups of experts recruited for answering the survey questions in 2009 and 2016. Another limitation of this data type is that we cannot say what types of policy change were taken into consideration by the

Table 10.1 Mexican standards for sulfur dioxide emissions from stationary sources

Year	Metropolitan area of Mexico City	Critical areas	Rest of the country
1988	–	Min: 1700 ppm Max: 5700 ppm	Min: 3400 ppm Max: 9500 ppm
1993	Min: 800 Max: 1100 ppm	Min: 1200 Max: 1650 ppm	Min: 2200 ppm 2600 ppm
1994	550 ppm	1100 ppm	2200 ppm

Source: Own elaboration based on Tosun (2008). The Mexican legal acts consulted are the following: NTE-CCAT-005/88 and NTE-CCAT-007/88, NOM-CCAT-019-ECOL/1993 (NE) and NOM-085-Semarnat-1994. PPM stands for parts per million.

respondents, if any. Nevertheless, survey data can be employed for analyzing policy change. Furthermore, it appears worth systematically comparing regulatory output data with survey data in order to see how strongly they correlate.

The last measurement approach to be discussed in this section refers to the extension of Hall's approach proposed by Knill, Schulze, and Tosun (2012). The authors argue that not only the setting but also the scope of a policy may be subject to change. To illustrate this point we look at the policy approach adopted by the Mexican government in order to prevent and control air emissions. In 1988, the government for the first time established maximum permissible limits for sulfur dioxide emissions from stationary sources. The standard defines limit values for so-called critical areas as well as the rest of the country. As we can infer from Table 10.1, in 1988, the Metropolitan Area of Mexico City was exempted from the regulatory standard, but this changed with the 1993 revision of the standard. Thus, in 1993, we can observe a change with regard to the scope of this particular policy instrument.

In sum, in this section, we showed how the original three dimensions of Hall's conceptualization as well as the modifications by Holzinger and Knill (2008) and Knill, Schulze, and Tosun (2012) allow for measuring different types of policy change. While the concept is most accurately operationalized by using policy output data, we showed that alternative data sources (e.g. survey data) are also compatible with it. With the first example already, we observed that the different dimensions are interdependent, which resulted in a nested data structure that has to be taken into account when analyzing the data. In practical terms, this multi-dimensional approach suffers from the need for a very careful coding of the data and therefore only one or few policies are examined. Even the ambitious approach by Holzinger, Knill, and Arts (2008) comprises an entirety of 40 environmental policies only.

4. THE BUDGET APPROACH TO MEASURING POLICY CHANGE

The desire to describe entire change patterns for one or several policy fields underlies the methodological approach of studies that build on the Punctuated Equilibrium Theory (Baumgartner and Jones 2009 [1993]). A prominent data type used by pertinent studies is changes in annual government spending data (e.g. Jones et al. 2009), but it is not the only data type used in this literature. Given the Punctuated Equilibrium Theory's interest in attention and agenda setting, media data (e.g. Van Aelst and Walgrave 2011) as well as statements in political documents

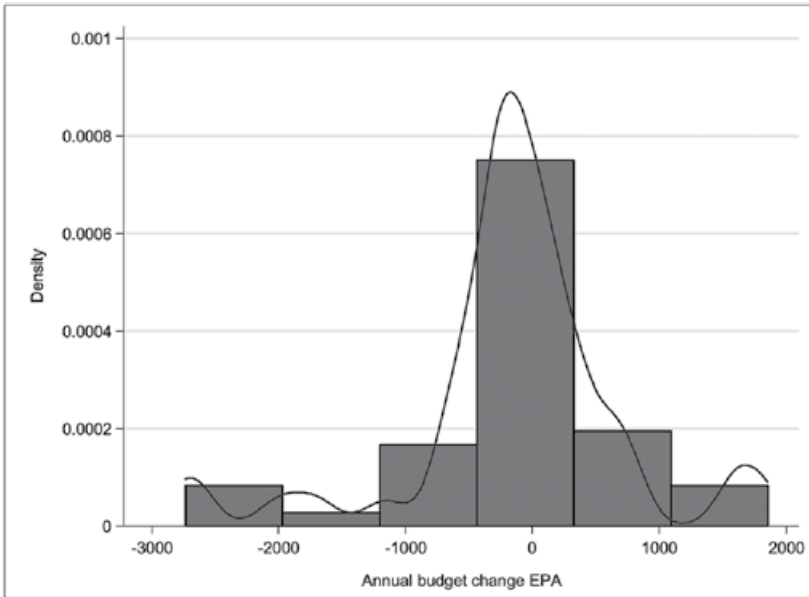
(e.g. Alexandrova et al. 2014; John and Bevan 2012) also represent important data types that are used to examine patterns of policy change. To facilitate comparative research on policy change, the Comparative Agenda Project developed a master codebook that guides the various data coding efforts (Bevan 2018). In this section, however, we concentrate on government budgets and how they change over time since we aim to introduce a different perspective on how we can measure policy change.

Government budgets can be assigned to a number of topics that correspond to policy domains, for which the annual percentage differences are calculated for each year. This produces positive, negative, or zero-values for the annual (percentage) differences. In its simplest form (see Breunig and Jones 2011), the data calculated is inspected by means of a graphical technique called a histogram. The histogram produced with the budget data is examined with regard to its mode value (i.e. the most frequent value), the skewness (i.e. symmetry), and the kurtosis (i.e. the degree to which the distribution is peaked). If the mode value is zero, there are many years for which annual percentage differences are zero, which equals to many instances of policy stability. If the distribution's left tail is longer and most of the distribution is at the right, it is left-skewed, whereas if the right tail is longer and most of the distribution is at the left, the distribution is right-skewed, which also provides important insights into whether the budget was mostly expanded or reduced over the observation period.

The kurtosis assesses whether the data are heavy-tailed (high kurtosis) or light-tailed (low kurtosis) relative to a normal distribution. A normal distribution has a kurtosis of 3 and is called *mesokurtic*, but the studies of changes in budgetary data are typically interested in kurtosis that is different from that of a normal distribution. A distribution with kurtosis greater than 3 (*platykurtic*) has shorter tails and a rather low and broad central peak compared to the normal distribution. A distribution with kurtosis smaller than 3 (*leptokurtic*) has longer tails, and the central peak is often found to be high and pointed. The latter is the type of distribution reported by studies drawing on the Punctuated Equilibrium Theory (see Breunig and Jones 2011).

To illustrate the descriptive statistics introduced above, we inspect Figure 10.4 that presents the annual changes in the budget of the United States Environment Protection Agency (EPA). In contrast to the classic applications of the budget approach, we do not examine a government budget for an entire policy (sub)field, but are interested in how the budget of one specific agency has changed over time. One may object that the EPA's budget might be too distant from actual policy making, but such objections can easily be dismissed. The EPA is extremely influential in reviewing the state of technology and monitoring environmental conditions, which places it in an important position to propose new standards and policies or recommend the existing ones to be reviewed. The lower the EPA's budget, the more it has to focus its activities on certain priority areas, which is likely to reduce the attention it pays to less pressing or important areas of environmental policy.

Figure 10.4 represents a histogram that uses data for the annual budget changes of the EPA 1970–2017. Added to the histogram is a Gaussian density function, which offers an even more abstract visualization of how the data is distributed. We can infer from the figure that the mode value of the distribution is zero as claimed by studies of government budgets (e.g. Breunig and Jones 2011; Jones et al. 2009). In other words, when analyzing the EPA's budget for a period of well beyond 40 years, we can state that the most frequent type of change was a zero-change, that is, stability.



Source: Own elaboration based on EPA data retrieved from the following website: <https://www.epa.gov/planandbudget/budget>. The numbers refer to billions of US dollars.

Figure 10.4 Changes in the annual budget of the US EPA, 1970–2017

Table 10.2 contains additional information that helps us to get a better sense of how the EPA’s budget developed over time. We can see that on average, for the 47 years observed, the annual budget of the agency was reduced by US\$150 billion with a median annual budget change of US\$128 billion. The greatest annual budget decrease amounted to US\$2,735 billion and the greatest increase to US\$1,859 billion, which explains the considerable variance in the data of US\$774,414 billion. The inter-quartile range suggests that the middle 50% of the distributions is around US\$540 billion. The negative value for the skewness indicates that the data is skewed to the left. The L-kurtosis score of .36 indicates that the distribution is highly leptokurtic, suggesting that the EPA’s budget both significantly expanded and shrank in some years.

Table 10.2 Descriptive statistics for the changes in the EPA’s annual budget

Statistics	Values
Mean	-150.0958
Median	-127.683
Variance	774413.6
Inter-quartile range	540.18
Skewness	-.7159329
L-kurtosis	0.359
Minimum	-2734.89
Maximum	1858.878
Number of budget years	47

Note: The numbers refer to billions of US dollars.

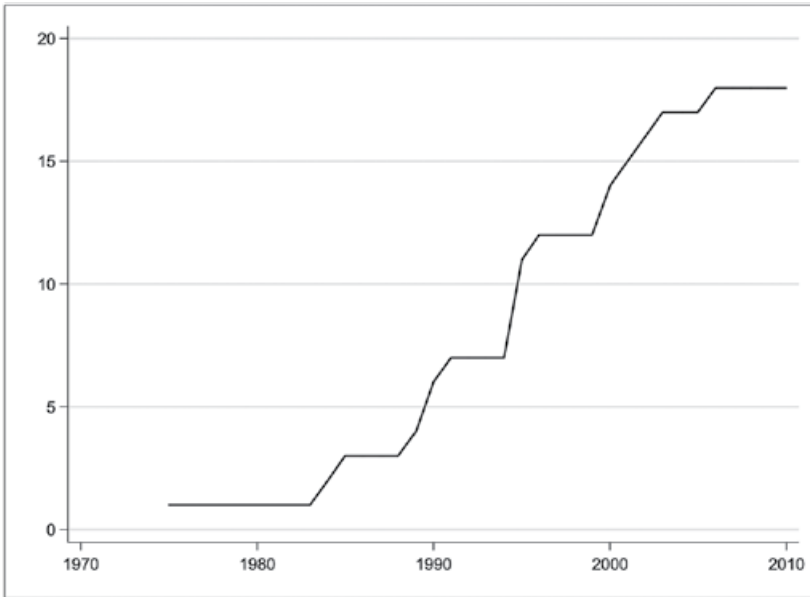
Taken together, the empirical impressions obtained by Figure 10.4 and Table 10.2 show that from 1970 to 2017, the EPA has most of the time experienced a stable budget, but at several instances it was also affected by drastic increases and decreases in its budget. These changes could have been induced by the ideological orientation of the US presidents and their administrations (e.g. Mendez 2018) as well as other political factors such as increased public demand for environmental protection as predicted by the Environmental Kuznets Curve (e.g. Van Alstine and Neumayer 2008) or macroeconomic factors such as the state of the economy. While the empirical illustration of this section referred to one specific government agency, it is possible to rely on budget data for identifying broader patterns of policy change. Bräuninger (2005), for example, examined both the levels and composition in government budgets and how these changed over time. This approach can be easily adopted for environmental subfields to obtain a broad yet differentiated empirical picture that allows for identifying government spending priorities and whether these shifted over time. For example, it is possible that spending on biodiversity has developed in a different fashion than spending on waste management and recycling.

While the budget approach has merit, it should be noted that it does not provide a measurement of policy making in the narrow sense. From that perspective, it can only be regarded as a proxy, but not as a direct measurement of policy change, especially since environmental policy is a regulatory policy where costs are predominantly imposed on the regulatees (Tosun 2012). Finally, a cautionary note on the availability of government budget data: in contrast to what one would expect, data availability is restricted and often the data quality is not very good. Therefore, researchers who are interested in adopting this approach to the measurement of policy change should bear in mind that they may have to invest heavily in obtaining the necessary data, especially if they are interested in data for a long observation period and for specific policy (sub)fields.

5. THE DIFFUSION APPROACH TO MEASURING POLICY CHANGE

The third approach to measuring policy change is about the first-time adoption of a policy innovation. In fact, we have already encountered this approach in the previous section in Figure 10.1 that showed in which year the countries adopted a policy to regulate the organic pollution of surface waters. Figure 10.5 is a modification of Figure 10.1 and presents the data as a diffusion curve, which is based on the cumulative number of policy adoption by 18 Central and South American countries between 1975 and 2010. We can observe the typical S-shaped curve that is indicative of diffusion processes (see Gray 1973). Diffusion data can be analyzed in two ways: First, we can assess the degree of policy change across these two regions; second, we can examine the individual countries and determine the factors inducing them to adopt a policy to regulate the organic pollution of surface waters.

Similar insights are provided by Figure 10.6, which is based on data on the diffusion of environmental ministries in the Americas as reported by Aklin and Urpelainen (2014), who build on Busch and Jörgens (2005a, 2005b). The dataset comprises an impressively long observation period of 50 years (1960–2010). This time we use a different methodology to visualize the diffusion of environmental ministries, which is known as the Kaplan-Meier failure function. The graph shows how the chances to experience a “failure”, i.e. the creation of a ministry, changes

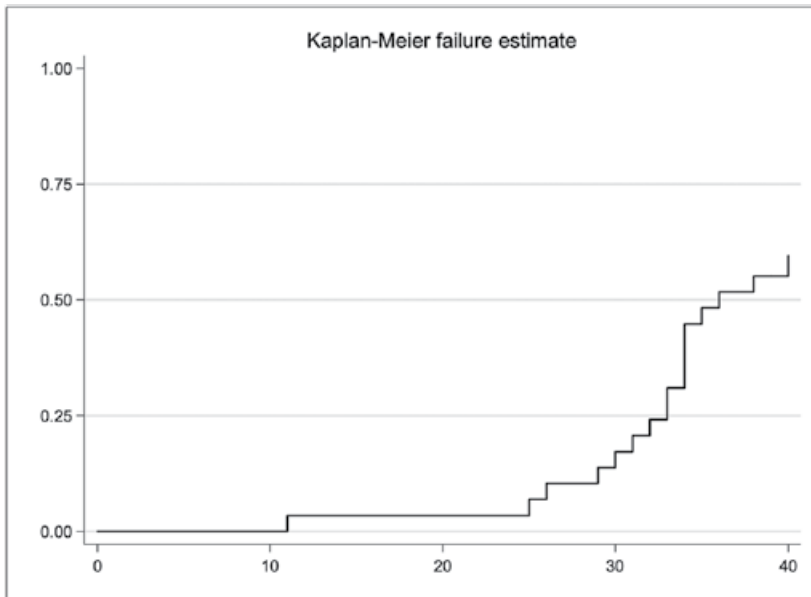


Source: Own elaboration based on Tosun (2013).

Figure 10.5 Diffusion of policies for regulation organic water pollution, 1975–2010

over time. At the beginning of the observation period, the chance of observing the creation of an environmental ministry is low, but it increases after about 30 years, that is, after the year 1990 and continues to do so until the year 2000. After that year, however, the countries that did not establish a ministry by then abstained from doing so, which is the reason why the values stagnate after that point in time.

In contrast to the former measurement approaches, the diffusion perspective captures one specific type of policy change only. It is best suited for examining the diffusion of policy paradigms, policy instruments, and policy instruments, but less so for policy settings and scopes. The main interest of this approach lies in the timing of policy change, which receives less attention with the other approaches. The methods for analyzing diffusion data are therefore primarily interested in explaining which factors accelerate or delay the adoption of a policy innovation by observation units. The data is typically coded as a dichotomous variable (0 = non-adoption; 1 = adoption). At the same time, this simplified way of measuring change makes this approach easy to implement since it can handle a lower level of empirical information. However, with the diffusion perspective, subtle changes cannot be captured and another (conceptual) question is how to cope with changes subsequent to the first-time adoption of a policy innovation (see Biesenbender and Tosun 2014), which are usually not taken into account. For a scholar of policy change, ending the data observation after the first-time adoption is often not desirable, which explains why policy change and policy diffusion are two phenomena studied by different research communities.



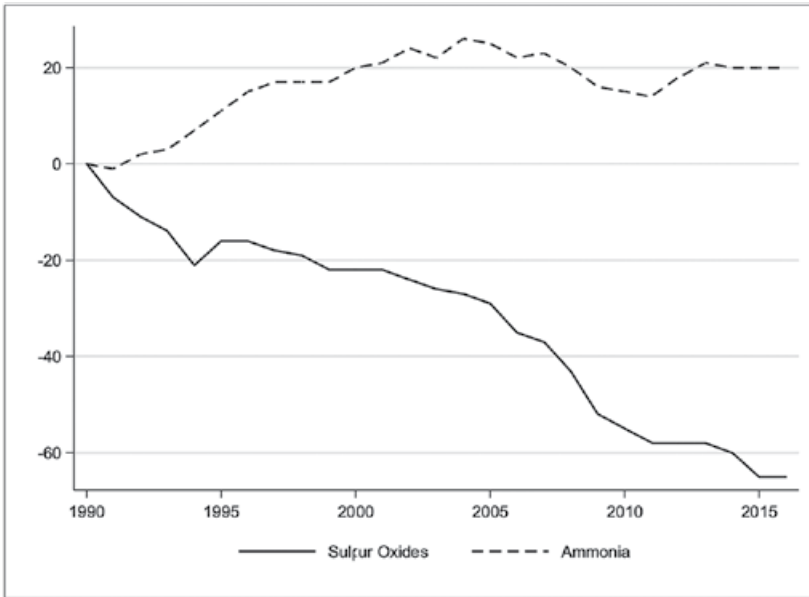
Source: Own elaboration based on Aklin and Urpelainen (2014).

Figure 10.6 *Diffusion of environmental ministries in the Americas, 1960–2010*

6. THE OUTCOME APPROACH TO MEASURING POLICY CHANGE

In the previous section, we presented measurements of policy change that originate from the conceptual and/or theoretical literature in comparative public policy. However, there also exist measurement approaches to (environmental) policy change that are rooted in different literatures such as comparative politics or political economy. Studies based in that literature also use outcome data as a proxy for measuring environmental policy change (e.g. Bernauer and Koubi 2009; Esty and Porter 2005; Neumayer 2003; Perkins and Neumayer 2008; Roller 2005). In the field of environmental policy, outcome data assesses levels of pollutant emissions or consumption rates for substances causing pollution (e.g. Wälti 2004).

Of course, policy making aims to change real-life conditions by means of adopting policies, but the relationship between a policy decision and its effects is affected by a considerable number of additional factors. One of these factors refers to whether a policy is implemented properly by the government agencies in charge of it and whether the target population of an environmental policy complies with the stipulations of the policy concerned (Tosun 2012). Seen from that perspective, outcome data – sometimes also referred to as performance data (e.g. Roller 2005) – potentially offer more information about the practical implications of policy change than policy outputs. However, the data that is gathered and provided by national agencies and international organization does not only contain implementation-related variation in the data, but also derives from many other sources that are difficult to identify



Source: Own elaboration based on data from the following website: <https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/air-pollutant-emissions.html>. (Accessed 1 November 2018).

Figure 10.7 Air pollutant emissions, Canada, 1990–2016 (percentage changes from 1990 level)

and therefore to control. Nonetheless, it can be worth consulting outcome data in order to get a basic sense of policy activities.

Figure 10.7 presents data on percentage changes of emissions of sulfur oxides and ammonia in Canada between 1990 and 2016. We can see that the sulfur oxide emissions went down rapidly after 1990, whereas ammonia emissions increased steadily. The main source of ammonia is agriculture, which also explains why the main emitters are the rural areas in Canada. The opposite development of these two emissions curves could be determined by different policy adopted, that is, no or lax standards for ammonia and strict standards for sulfur oxide. However, this we cannot directly read out of the outcome data and we need additional data on policy outputs to be able to answer this question. But what this figure certainly presents is an empirical puzzle that can guide an analysis of policy change based on policy output measures.

In practical terms, this measurement has two advantages: first, the availability of outcome data; second, outcome data are often accessible in form of metric variables so that even small differences between the countries and changes over time become apparent. Turning to the limitations, even though data availability is high for most democratic countries, with some outcome indicators, there is a lack of data for developing and transition countries. Another issue concerns the validity of the data provided by countries, especially the non-democratic ones. More generally, when using policy outcome data as a proxy for policy change, we need to take into account that shifts in policy outcomes cannot reliably be traced back to changes in policy outputs (see Knill, Schulze, and Tosun 2012). Instead, we have to examine whether

additional variables may confound the measurement (e.g. Stern 2004). Once potentially interfering variables can be excluded or are under control, environmental outcome data can principally be used as an indicator for policy change.

7. DISCUSSION AND CONCLUSION

In this chapter, we gave an overview of ways to measure policy change. The public policy literature is characterized by a multitude of concepts that all somehow address (environmental) policy change. Nevertheless, existing studies of policy change do not represent a coherent body of research that proposes a generally applicable concept of policy change. While one may initially expect this to produce a fragmented literature, the range of measurements available shows how prolific this research is. Furthermore, the measurement concepts presented in this chapter are truly complementary. With the multi-dimensional approach that builds on Hall's (1993) work, we can obtain a refinement measurement of policy change. With the government budget approach that is often associated with the Punctuated Equilibrium Theory (Baumgartner and Jones 2009 [1993]), we can assess patterns of policy change. The diffusion approach allows for concentrating on the first-time adoption of a policy innovation. The complementarity of these approaches also became clear in this chapter when we used the same data source for illustrating changes in policy presence and policy presence.

Which measurement approach one chooses should depend on the research interest and the data availability, which is an important issue and often entails the need to collect and code data. In this context, we could also observe that the dimensions identified by Hall (1993) are not independent from each other, which makes the coding process even more demanding (e.g. Holzinger, Knill, and Arts 2008). More simple measurements are offered by government budget data and data on the diffusion of policy innovations, but these might be too limited concerning their level of empirical information to address the respective research questions. Generally, the most demanding form of policy change to measure concerns that of policy paradigms since it requires information about the ideas underlying policies (e.g. Fontaine, Narvaez, and Velasco 2018). Therefore, it is plausible that most studies interested in changes in policy paradigms adopt qualitative approaches. Nevertheless, it should be noted that there exist numerous studies that apply quantitative methods for studying policy change. Among these studies, the greatest share relies on the Punctuated Equilibrium Theories (e.g. Baumgartner et al. 2009).

In sum, we can show that the literature has progressed well in developing a set of different measurement approaches, which are also associated with theoretical concepts. When consulting the pertinent literature, on the one hand, we find a great amount of conceptual discussions of policy change that do not engage with measurement issues. On the other hand, we have many empirical studies that do not reflect on the conceptual basis of measuring policy change. We invite future research to make an effort in aligning the conceptual discussions with empirical work (e.g. Peters, Fontaine, and Mendez 2018). Furthermore, we encourage using different measurement approaches and comparing their outcomes with one another. Thus, there are still many methodological issues that merit attention from scholars in comparative public policy.

REFERENCES

- Aklin, M. and Urpelainen, J. (2014). The global spread of environmental ministries: domestic–international interactions. *International Studies Quarterly* 58 (4): 764–80.
- Alexandrova, P., Carammia, M., Princen, S., and Timmermans, A. (2014). Measuring the European Council agenda: introducing a new approach and dataset. *European Union Politics* 15 (1): 152–67.
- Baumgartner, F. R., Berry, J. M., Hojnacki, M., Leech, B., and Kimball, D. C. (2009). *Lobbying and Policy Change: Who Wins, Who Loses, and Why*. Chicago, IL: University of Chicago Press.
- Baumgartner, F. R. and Jones, B. D. (2009 [1993]). *Punctuated Equilibria in Politics: Agendas and Instability in American Politics*. Chicago, IL: University of Chicago Press.
- Béland, D. (2009). Ideas, institutions, and policy change. *Journal of European Public Policy* 16 (5): 701–18.
- Bernauer, T. and Koubi, V. (2009). Effects of political institutions on air quality. *Ecological Economics* 68 (5): 1355–65.
- Bevan, S. (2018). *Gone Fishing: The Creation of the Comparative Agendas Project Master Codebook*. Edinburgh: University of Edinburgh.
- Biesenbender, S. and Tosun, J. (2014). Domestic politics and the diffusion of international policy innovations: how does accommodation happen? *Global Environmental Change* 29: 424–33.
- Bräuninger, T. (2005). A partisan model of government expenditure. *Public Choice* 125 (3–4): 409–29.
- Breunig, C. and Jones, B. D. (2011). Stochastic process methods with an application to budgetary data. *Political Analysis* 19 (1): 103–17.
- Browne, C., Di Batista, A., Geiger, T., and Verin, S. (2016). The executive opinion survey: the voice of the business community. In K. Schwab (Ed.), *The Global Competitiveness Report 2016–2017*. Geneva: World Economic Forum, 77–88.
- Browne, C. and Geiger, T. (2009). The executive opinion survey: capturing the views of the business community. In X. Sala-i-Martin (Ed.), *The Global Competitiveness Report 2009–2010*. Geneva: World Economic Forum, 49–58.
- Busch, P. and Jörgens, H. (2005a). International patterns of environmental policy change and convergence. *European Environment* 15 (1): 80–101.
- Busch, P. and Jörgens, H. (2005b). The international sources of policy convergence: explaining the spread of environmental policy innovations. *Journal of European Public Policy* 12 (5): 1–25.
- Cairney, P. and Jones, M. D. (2016). Kingdon’s multiple streams approach: what is the empirical impact of this universal theory? *Policy Studies Journal* 44 (1): 37–58.
- Capano, G. (2009). Understanding policy change as an epistemological and theoretical problem. *Journal of Comparative Policy Analysis* 11 (1): 7–31.
- Dobbin, F., Simmons, B., and Garrett, G. (2007). The global diffusion of public policies: social construction, coercion, competition, or learning? *Annual Review of Sociology* 33: 449–72.
- Dunlop, C. A. and Radaelli, C. M. (2013). Systematising policy learning: from monolith to dimensions. *Political Studies* 61 (3): 599–619.
- Esty, D. C. and Porter, M. E. (2005). National environmental performance: an empirical analysis of policy results and determinants. *Environment and Development Economics* 10 (4): 391–434.
- Fontaine, G., Narvaez, I., and Velasco, S. (2018). Explaining a policy paradigm shift: a comparison of resource nationalism in Bolivia and Peru. *Journal of Comparative Policy Analysis: Research and Practice* 20 (2): 142–57.
- Geva-May, I. (2004). Riding the wave of opportunity: termination in public policy. *Journal of Public Administration Research and Theory* 14 (3): 309–33.
- Gray, V. (1973). Innovation in the states: a diffusion study. *American Political Science Review* 67 (4): 1174–85.
- Hagopian, F. and Mainwaring, S. P. (Eds.) (2005). *The Third Wave of Democratization in Latin America: Advances and Setbacks*. New York: Cambridge University Press.
- Hall, P. A. (1993). Policy paradigms, social learning, and the state: the case of economic policymaking in Britain. *Comparative Politics* 25 (3): 275–96.
- Hogwood, B. W. and Peters, B. G. (1982). The dynamics of policy change: policy succession. *Policy Sciences* 14 (3): 225–45.

- Holzinger, K. and Knill, C. (2008). Theoretical framework: causal factors and convergence expectations. In K. Holzinger, C. Knill, and B. Arts (Eds.), *Environmental Policy Convergence in Europe? The Impact of International Institutions and Trade*. Cambridge: Cambridge University Press, 30–63.
- Holzinger, K., Knill, C., and Arts, B. (Eds.) (2008). *Environmental Policy Convergence in Europe? The Impact of International Institutions and Trade*. Cambridge: Cambridge University Press.
- Howlett, M. and Cashore, B. (2009). The dependent variable problem in the study of policy change: understanding policy change as a methodological problem. *Journal of Comparative Policy Analysis* 11 (1): 33–46.
- Howlett, M. and del Rio, P. (2015). The parameters of policy portfolios: verticality and horizontality in design spaces and their consequences for policy mix formulation. *Environment and Planning C: Government and Policy* 33 (5): 1233–45.
- Howlett, M. and Ramesh, M. (2002). The policy effects of internationalization: a subsystem adjustment analysis of policy change. *Journal of Comparative Policy Analysis* 4 (3): 31–50.
- Jensen, C., Knill, C., Schulze, K., and Tosun, J. (2014). Giving less by doing more? Dynamics of social policy expansion and dismantling in 18 OECD countries. *Journal of European Public Policy* 21 (4): 528–48.
- Jepson, W. E., Brannstrom, C., and Stancato de Souza, R. (2008). Brazilian biotechnology governance: consensus and conflict over genetically modified crops. In G. Otero (Ed.), *Food for the Few: Neoliberal Globalism and Biotechnology in Latin America*. Austin, TX: University of Texas Press, 217–42.
- John, P. and Bevan, S. (2012). What are policy punctuations? Large changes in the legislative agenda of the UK government, 1911–2008. *Policy Studies Journal* 40 (1): 89–108.
- Jones, B. D., Baumgartner, F. R., Breunig, C., Wlezien, C., Soroka, S., Foucault, M., Abel, F., Green-Pedersen, C., Koski, C., John, P., Mortensen, P. B., Varone, F., and Walgrave, S. (2009). A general empirical law of public budgets: a comparative analysis. *American Journal of Political Science* 53 (4): 855–73.
- Jordan, A., Bauer, M. W., and Green-Pedersen, C. (2013). Policy dismantling. *Journal of European Public Policy* 20 (5): 795–805.
- Kingdon, J. W. (2002). *Agendas, Alternatives, and Public Policies*. New York: Longman.
- Knill, C., Schulze, K., and Tosun, J. (2012). Regulatory policy outputs and impacts: exploring a complex relationship. *Regulation & Governance* 6 (4): 427–44.
- Knill, C., Tosun, J., and Heichel, S. (2008). Balancing competitiveness and conditionality: environmental policy-making in low-regulating countries. *Journal of European Public Policy* 15 (7): 1019–40.
- Kochtcheeva, L. V. (2009). *Comparative Environmental Regulation in the United States and Russia: Institutions, Flexible Instruments, and Governance*. Albany, NY: SUNY Press.
- Maggetti, M. and Gilardi, F. (2016). Problems (and solutions) in the measurement of policy diffusion mechanisms. *Journal of Public Policy* 36 (1): 87–107.
- Maor, M., Tosun, J., and Jordan, A. (2017). Proportionate and disproportionate policy responses to climate change: core concepts and empirical applications. *Journal of Environmental Policy & Planning* 19 (6): 599–611.
- Mendez, J. L. (2018). Presidents' strategic action and policy change in Mexico: a comparison of the hydrocarbon policy reforms of Enrique Peña and Felipe Calderon. *Journal of Comparative Policy Analysis: Research and Practice* 20 (2): 209–22.
- Neumayer, E. (2003). Are left-wing party strength and corporatism good for the environment? Evidence from panel analysis of air pollution in OECD countries. *Ecological Economics* 45 (2): 203–20.
- Nohrstedt, D. and Weible, C. M. (2010). The logic of policy change after crisis: proximity and subsystem interaction. *Risk, Hazards & Crisis in Public Policy* 1 (2): 1–32.
- OECD (2016). *Environmental Performance Reviews: Peru*. Paris: OECD.
- Perkins, R. and Neumayer, E. (2008). Fostering environment efficiency through transnational linkages? Trajectories of CO₂ and SO₂, 1980–2000. *Environment and Planning A* 40 (12): 2970–2989.
- Peters, B. G., Fontaine, G., and Mendez, J. L. (2018). Substance and methods in the comparative study of policy change. *Journal of Comparative Policy Analysis: Research and Practice* 20 (2): 133–41.
- Peters, B. G., Jordan, A., and Tosun, J. (2017). Over-reaction and under-reaction in climate policy: an institutional analysis. *Journal of Environmental Policy & Planning* 19 (6): 612–24.

- Pralle, S. B. (2006). Timing and sequence in agenda-setting and policy change: a comparative study of lawn care pesticide politics in Canada and the US. *Journal of European Public Policy* 13 (7): 987–1005.
- Roller, E. (2005). *The Performance of Democracies: Political Institutions and Public Policy*. Oxford: Oxford University Press.
- Schneider, A. and Ingram, H. (1993). Social construction of target populations: implications for politics and policy. *American Political Science Review* 87 (2): 334–7.
- Stern, D. I. (2004). The rise and fall of the environmental Kuznets curve. *World Development* 32 (8): 1419–39.
- Tosun, J. (2008). *Environmental Policy in Chile and Mexico: Explaining the Effect of Economic Integration on Regulatory Standards*. Saarbrücken: VDM.
- Tosun, J. (2012). Environmental monitoring and enforcement in Europe: a review of empirical research. *Environmental Policy and Governance* 22 (6): 437–48.
- Tosun, J. (2013). *Environmental Policy Change in Emerging Market Democracies: Central and Eastern Europe and Latin America Compared*. Toronto: University of Toronto Press.
- Tosun, J. (2018). Diffusion: an outcome of and an opportunity for polycentric activity? In A. Jordan, D. Huitema, H. van Asselt, and J. Forster (Eds.), *Governing Climate Change: Polycentricity in Action*. Cambridge: Cambridge University Press, 152–68.
- Tosun, J. and Workman, S. (2017). Struggle and triumph in fusing policy process and comparative research. In C. Weible and P. Sabatier (Eds.), *Theories of the Policy Process*. Boulder, CO: Westview Press, 329–62.
- Van Aelst, P. and Walgrave, S. (2011). Minimal or massive? The political agenda-setting power of the mass media according to different methods. *The International Journal of Press/Politics* 16 (3): 295–313.
- Van Alstine, J. and Neumayer, E. (2008). The environmental Kuznets curve. In K. P. Gallagher (Ed.), *Handbook on Trade and the Environment*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing, 49–59.
- Vogel, D. (1986). *National Styles of Regulation: Environmental Policy in Great Britain and the United States*. Ithaca, NY: Cornell University Press.
- Walker, J. L. (1969). The diffusion of innovations among the American states. *American Political Science Review* 63 (3): 880–899.
- Wälti, S. (2004). How multilevel structures affect environmental policy in industrialized countries. *European Journal of Political Research* 43 (4): 599–634.
- Weible, C. M. and Sabatier, P. (Eds.) (2017). *Theories of the Policy Process*. Boulder, CO: Westview Press.

11. Using indexes in comparative policy analysis: global comparisons

Tero Erkkilä

1. INTRODUCTION

Global indicators have become influential as policy instruments (Erkkilä 2013; Espeland and Sauder 2007; Kelley and Simmons 2015; Merry, Davis, and Kingsbury 2015), effecting national policies and serving as assessment criteria when making decisions on countries' eligibility for loans and development funding. This has also raised interest in the composition of these measurements and their methodology, as well as their underlying ideological and political aspects (Erkkilä and Piironen 2009; Löwenheim 2008). Global country rankings and indicators have faced criticism for their normative character and methodology. In particular, the global rankings have been criticized for using blunt aggregate figures in establishing rank orders. Because of this, there have been attempts at creating so-called second-generation governance indicators that provide more detailed and reform-oriented measurements of governance.

My chapter analyzes measurement issues that arise in developing indices for policy in an international context. Attention is given to the rise of second-generation governance indicators and related methodological debates and changes in the production and use of indicators. I use measurements of transparency as a case for analyzing the above shift towards second-generation governance indicators. However, at the same time, there has been a recent surge of regional and city-level rankings of competitiveness and innovation. Though cities are becoming increasingly important in public policy, the increased interest in measuring policies and institutional practices on city and regional level can also be explained by the global knowledge brokers' interest to enter the ranking activity. Aiming to provide "holistic" measurements, the sub-national indicator sets also contain variables to measure good governance.

Here the problems identified in the early rankings of good governance are again encountered. I argue that these measurement issues can be understood against the field development in global ranking, where the actors hoping to enter the field tend to reproduce existing practices in the field, sharing normative beliefs and data. Sartori famously stated that comparative analyses "travel on their concepts" (Sartori 1970). The current comparative assessments in policy indicators rather seem to travel on availability of existing data than on concepts. The new regional and city-level measurements are mostly composite indicators, drawing their data from various sources, such as existing national-level data sets already existing. This leads to conceptual stretching and other measurement issues.

2. BACKGROUND

Since their emergence in 1990s, governance indices have become a central policy instrument of transnational governance. Even though there is no apparent mechanism of influence, the governance indices are said to have a strong steering effect through “naming and shaming” on the countries that are being ranked, as policy actors adhere to the perceived norm (Buduru and Pal 2010; Davis, Kingsbury, and Merry 2012; Erkkilä and Piironen 2009; Löwenheim 2008; Merry 2011). Some observers have likened this reflexivity to Foucauldian governance from a distance or the Weberian iron cage (Erkkilä and Piironen 2009; Löwenheim 2008; Miller and Rose 1990). Scholars have also described policy indicators as potential tools of “social pressure” (Kelley and Simmons 2015), “reactivity” (Espeland and Sauder 2007), and “quiet power” (Merry, Davis, and Kingsbury 2015) with “indirect policy effects” (Gornitzka 2013). Moreover, such policy measures also have the potential for unintended consequences (Espeland and Sauder 2007; Pidd 2005; Robinson 2003; Smith 1995; Thiel and Leeuw 2002): numbers matter, but not always in the way designed.

Numbers seem neutral, but indicators just like any statistics are fundamentally political (Desrosières 1998; Porter 1996). They often carry ideological baggage owing to the choices over concepts and their attributes, data sources, and methodology, such as weighting of data. Even the presentation of data may have political implications, as aggregating results for creating rank orders is likely to create headlines, whereas disaggregated presentation of data is likely to speak to expert audiences only. Furthermore, it is important to note that also bad policy indicators can and do have policy effects. This has drawn interest in the composition of these measurements and their methodology.

Recently, there has been a shift in the way governance is assessed globally, as more nuanced and detailed numerical assessments, often referred to as second-generation or actionable indicators, are challenging rankings (Knack, Kugler, and Manning 2003; Trapnell 2011). Second-generation indicators are characterized by (1) transparency, meaning that they should be replicable, well-documented, and non-controversial; (2) availability of data, meaning broad country coverage and continuity over time; (3) quality and accuracy as well as validity of measurements; and (4) specificity, meaning that indicators measure specific institutions or output (Knack, Kugler, and Manning 2003). Second-generation indicators are also often referred to as “mappings”, as they allow different (disaggregated) representation of data, instead of just a single aggregate number.¹

We can understand this shift in the production of governance indicators as field structuration, where there are new actors joining the field of global governance assessments with competing sets of indicators (Kauppi and Erkkilä 2011). The developments in the field not only demonstrate collaboration between the knowledge producers but also competition that becomes apparent only sporadically. The index producers share common normative and causal perceptions as well as data sources (Erkkilä and Piironen 2009; Haas 1992), but also compete with each other for visibility and users (Kauppi and Erkkilä 2011). The indicators are also policy instruments, functioning as means for collecting information but also as effective tools of government in trying to influence the outside world (cf. Hood and Margetts 2007: 3). Hence, changes in their outlook have implications for their mechanisms of influence.

Table 11.1 summarizes developments in the field. The first global indicators tend all to be rankings, i.e. they are based on single aggregate figures that allow making rank orders on the subjects of measurement. Such data sets were initially produced by international organiza-

Table 11.1 *Rankings and second-generation governance indicators (after Erkkilä 2016)*

	Rankings	Second-generation governance indicators
Presentation and specificity of results	Single aggregate figure: general systemic-level information	Disaggregated data: detailed information on institutions and processes
Production of data	In collaboration and competition between international organizations, but also NGOs and private companies	Mostly by NGOs but also international organizations in intensive collaboration between data producers
Use and mechanism of influence	General-level assessments, comparisons, naming and shaming, adherence to norm	Monitoring and expert knowledge, peer pressure, funding

tions, though some NGOs and private organizations have also been involved in this. Rankings allow general-level assessments and comparisons, but also naming and shaming, pressuring national governments and policy makers to adhere to what seems to be a global norm. More recently, there has been a shift towards so-called second-generation governance indicators, whose presentation is disaggregated and hence not suited for making ranking lists.

This is also intentional, as the second-generation indicators aim to provide more detailed information on institutions and processes, also aiming for their active development. They allow collecting detailed expert knowledge mainly used for monitoring activities. Their mechanism of influence is hence not public naming and shaming, but peer pressure between ranked entities. Disaggregated indicators can be used for funding decisions (as can be rankings) and they have also acquired such uses. The second-generation measurements are mostly produced by smaller NGOs in intensive collaboration between data producers. But while the qualitative changes in the rankings are limited, there are significant changes in the uses and mechanisms of influence of governance indicators.

In the following, I will first explore the shift from rankings to second-generation governance indicators. I will focus on measurements of transparency, which is widely acknowledged to be a key component of good governance. Nevertheless, few governance indices or country rankings have actually measured transparency in the past, an indication of its poor measurability. Because of the shift from rankings to second-generation governance indicators, transparency has also become a measurable attribute of good governance assessments.² This also marks a turn away from aggregate rankings towards the more detailed and customized presentation of data. But most importantly there is an attempt to have conceptually more focused analysis of specific institutional aspects of governance.

Though conceptual changes in measurements remain limited, the field is moving towards more advanced measurements of good governance. However, new regional and city-level measurements of competitiveness and innovation have emerged at the same time. These are mostly composite indicators, drawing their data from various sources. They now largely contain all the methodological problems identified earlier in the good governance measurements. At present, the global comparative assessments do not “travel on their concepts” (Sartori 1970), but on availability of existing data. This now leads to conceptual stretching, but is also related to various other measurements issues that prevail despite attempts for establishing criteria for second-generation indicators. I link these measurement issues to the field development and process of producing global indicators.

3. GOVERNANCE INDICATORS AND CRITIQUE OF RANKING

Global governance indicators differ from the previous social scientific attempts at comparing countries (Erkkilä and Piironen 2009). Whereas large country comparisons were previously done mostly by academics, the practice today is largely in the hands of international governmental organizations (the World Bank, United Nations Development Programme [UNDP], Organisation for Economic Co-operation and Development [OECD]), non-governmental organizations (NGOs) (such as Transparency International, Freedom House), private businesses (such as Standard & Poor's) and linked associations (including the World Economic Forum [WEF] and Bertelsmann Foundation).

Also, the subject of measurement has shifted. Whereas previous academic assessments centered on the notion of democracy, the current country comparisons focus on good governance (Erkkilä and Piironen 2009: 130–132). Many assessments root the notion of “good governance” in efficiency-seeking perceptions of institutions put forward by international organizations of economic development (Drechsler 2004; Seppänen 2003; Zanotti 2005).

The rise of the governance indices is related to the global concern over good governance and corruption (Ivanov 2009). In coinciding with the general pressures for economic globalization, the concern over good governance led to the development of governance indicators. The first of its kind was the World Bank Institute's Worldwide Governance Indicators (WGI). The WGI initially targeted specific problems of global governance, such as corruption. But arguably the different measurements of corruption and accountability were not always coherent in their results and the developers of the WGI tried to neutralize this variance by creating an aggregate number of the available measurements (Erkkilä 2016). As a composite indicator the WGI has become a model for various data sets to follow. The somewhat unanticipated effect of aggregation was that it gave the figures high media visibility when the aggregated results were presented as rank orders.

Aggregation allows ranking nations based on their relative position on the various measurements and the league table format has drawn a fair amount of media attention on certain global measurements. At the same time, the rankings have faced criticism. Most notably, rankings have been controversial in countries that fare poorly in them, indicating the politicization of governance indicators (Guha and McGregor 2007; Harding 2013). The most visible critique of the rankings has been methodological, sparking a lively debate with and among the developers (Kaufmann, Kraay, and Mastruzzi 2010, 2011; Thomas 2010). The criticism of the existing rankings – WGI in particular – has led to attempts to develop indicators that are more appropriate and methodologically advanced (Andrews, Hay, and Myers 2010; Gramatikov, Barendrecht, and Verdonschot 2011; Joshi 2011; McFerson 2009).

Aggregation, the aim to make single ranking numbers, has drawn much attention to the first generation of governance indicators (Langbein and Knack 2010). While the conceptualization of good governance, which is inclined towards market liberalism, has not been a broadly politicized topic, aggregation as a methodological choice has become a subject of political controversy. The WGI have been criticized for the use of aggregate figures by the OECD, which has been creating its own non-aggregated data set. Also the development team of the WGI has downplayed its optimism about aggregation over time (Erkkilä and Piironen 2014), and ultimately denounced ranking as a technique for comparing individual countries (Kaufmann, Kraay, and Mastruzzi 2008: 5). Another methodological debate addresses the validity of the

measurements and the measurability of abstract issues (Andrews 2008; Barendrecht 2011; Ginsburg 2011; Neumann and Graeff 2010; van de Walle 2006). Moreover, the global indices might not always be apt for observing grassroots developments, and might even overlook undemocratic developments or crises of governance (Hinthorne 2011; Morgan 2011).

The criticism of rankings has led to a shift in the global field of governance assessment towards “second-generation” or “actionable” (good) governance indicators. This is explored below through the case of transparency metrics.

4. TOWARDS SECOND-GENERATION MEASUREMENTS: TRANSPARENCY METRICS

Following the critique of ranking, global governance measurements have evolved towards second-generation governance indicators, becoming more issue-specific measurements, as is apparent in the methodological development of transparency metrics. The development towards more “actionable” governance indicators has drawn attention to particular aspects of governance, instead of the ranking of countries. As the name indicates, the organizations behind actionable indicators are actively attempting to influence and improve the measured aspects of governance (Trapnell 2011). Whereas the early rankings did not necessarily measure transparency, the new second-generation rankings increasingly include various concrete assessments of it. The shift can be observed from Table 11.2 showing the development of the global measurements of good governance with regard to the assessments of transparency. There are also new actors on the scene, as various organizations are entering global comparative assessment.

The first rankings of good governance largely built on the ideas of the so-called “Washington consensus”, a notion that comprised the key causal beliefs of economic development of past decades, including the idea of avoiding information asymmetries in the market through transparency (Stiglitz 2002). Still, only a few of the early governance indices specifically measure transparency. One reason for this has been that transparency is very difficult to operationalize and measure (Hazell and Worthy 2010; Michener 2015). Another reason for the absence of measurements on transparency has been the predominance of aggregate figures in the early governance indicators, leading to assessments that are highly abstract.

The first ranking to assess transparency in some sense has been the Freedom in the World ranking by the Freedom House. Published already in 1973, the ranking assesses liberal democracy, concentrating mainly on civil liberties and political rights. This has involved assessment of government openness and transparency. Then came another Freedom House ranking produced in 1980 – Freedom of the Press – that assesses transparency as an element of media environment, with regard to legal, political, and economic conditions. These two rankings were for a long time the only global rankings to address the issue of transparency, even measuring it.

Subsequently, two prominent rankings of good governance emerged in the mid-1990s, the Corruption Perception Index (CPI) by Transparency International (in 1995) and the WGI by the World Bank (launched in 1996).³ These two rankings are related to the general rise of good governance as a global concern in the development economics which coincided with the paradigm shift in economics that stressed the role of information in the functioning of markets

Table 11.2 Global measurements of good governance and transparency

1970	1980	1990	2000	2010
Freedom in the World (1972)	Freedom of the Press (1980)	Corruption Perception Index (CPI) (1995)	Fringe Special (2001)	Global Integrity Report – Integrity Scorecard (2010)
		Worldwide Governance Indicators (WGI) (1996)	UN E-Government Readiness/ Development Index (EGRI/EGDI) (2003)	Global Right to Information (RTI) Rating (2011)
			UN E-Participation Index (2003)	Implementation Assessment Tool (IAT) (2011)
			Global Integrity Report – Global Integrity Index (2006)	
			Open Budget Index (2006)	
			Open Net Initiative (2007)	
			Actionable Governance Indicators – Public Accountability Measures (2008)	
			Government at a Glance (GG) (2009)	

(Stiglitz 1998, 2002). However, neither of these rankings directly measures transparency. Instead, they refer to transparency as an explanation for the results of the rankings.

Several other rankings that assessed transparency also emerged in the early 2000s. In 2002, the NGO Reporters without Borders launched its Press Freedom Index, which ranks countries on press freedom. The United Nations e-Government survey produced two rankings in the early 2000s, the E-Government Readiness Index and E-Participation Index that measure the accessibility of government data and legislative information as well as the participatory aspects of online transparency. As part of the global drive for anti-corruption (Ivanov 2009), Global Integrity launched its Global Integrity Index in 2006 that assessed governments' anti-corruption and accountability mechanisms.

However, because of the mounting criticisms of the use of rankings, there is a growing demand for non-aggregate figures, which have also influenced the way transparency is measured. Rankings to measure good governance and transparency have been complemented and challenged by non-aggregate, second-generation governance indicators. This development has also caused shifts in the activities of established index producers. A concrete example of this was Global Integrity's 2010 discontinuation of the Global Integrity Index, a widely cited ranking. Instead, Global Integrity now publishes its annual *Global Integrity Report* with an Integrity Scorecard, which maps selected aspects of government integrity. The organization itself referred to two reasons for the shift from ranking to mapping (Global Integrity 2011). First, the number of countries ranked by Global Integrity had diminished over time, thereby losing the assessment's global element. Second, the rankings were seen as too blunt in their assessment of governance and bringing no visible effects. This reflects the shift in the field of governance indices, away from the use of single aggregated figures towards more actionable and reform-oriented measurements.

Already in 2001, Roger Vleugel launched his *Fringe Special* initiative that compares freedom of information laws around the globe. While there have been similar initiatives (for example Freedominfo.org) the *Fringe Special* is perhaps the most up-to-date assessment of freedom of information acts, collected by a network of Dutch and foreign journalists. Fringe does not produce a ranking and the *Fringe Special* could be seen as an early form of

“mapping” that has come to complement the good governance rankings. Because of the shift towards actionability, the Fringe listing has gained new users recently though it was launched already a decade ago.

The transparency of finances has been a topical issue in good governance debates, and the International Budget Partnership has been collecting an Open Budget Index since 2006. This basically explores the public access to budget information (cf. Seifert, Carlitz, and Mondo 2013). Since 2007, the Open Net Initiative has produced a mapping of government censorship and filtration of the internet. We can also see the initiative as a representative of the second-generation governance indices that are critical about ranking countries, and instead make non-aggregate measurements of more specific qualities of governance.

Even large organizations of government reform such as the OECD have jumped the bandwagon by producing second-generation indicators on government performance. The OECD’s Government at a Glance (GG, launched in 2009) is more sophisticated than the rankings of governance performance or competitiveness (GCI, WGI, CPI), as it aims for a multi-dimensional assessment. As a newcomer to the production of governance indices, the OECD has argued strongly for the need of this new knowledge product on the basis that as a non-aggregate figure it marks a methodological improvement to the existing rankings, most notably to the WGI (OECD 2006: 7, 60; 2007: 3).⁴

Also, the World Bank has developed a systematic response to the methodologically and politically motivated criticism of the WGI. The World Bank’s governance and anti-corruption strategy of 2007 endorses the use of “disaggregated and actionable indicators” (World Bank 2007: ix). Related to this, the World Bank has developed a set of indicators, named Actionable Governance Indicators (AGI) alongside its WGI. This new set of indicators is reform-oriented and more nuanced, striving for close observations on selected issues of governance (Trapnell 2011).

It is important to note that the key actors not only collaborate in conceptual terms but also share data. This has been the case with the first rankings, such as the WGI, GCI, and CPI, but the “second-generation” or “actionable” governance indicators also build on the same logic. Most notably, the World Bank’s AGIs are fundamentally a collection of data produced by others. They aim to provide single access points to different indices that measure governance on a global scale. The AGIs also contain their own Public Accountability Measures component that assesses the legal framework of freedom of information, coverage of information, and procedures of accessing information. This aims at assessing not only legislation and reported procedures but also their implementation and institutional practices.

According to its developers, the AGIs were launched not to compete with the WGI but to complement it, though the AGIs entered the scene as criticism towards WGI was gaining ground, both within the Bank and outside. On a general level, the AGIs utilize the growing supply of governance indices that is mostly of a global nature. There are at present many small NGOs producing detailed measurements of certain aspects of governance that are becoming legitimate sources of information for large international organization such as the World Bank.

The Global RTI Rating by the Center for Law and Democracy and the Carter Center’s Implementation Assessment Tool (IAT), are representatives of this new development. Launched in 2011, they measure the right to information from a legal perspective (Global RTI Rating) and government compliance (IAT). These two indicators are clearly part of the new second-generation governance indices that have opened the way for smaller actors in the field. A noteworthy aspect of the new actors’ outlook is their North American origin (Global

Integrity, Open Net Initiative, Center for Law and Democracy, Carter Center). Concerning the normative and causal beliefs, they all seem to follow the key ideas of good governance and the Washington consensus, where transparency is a tool for both democracy and efficiency.⁵

The new index producers characteristically collect data through collective effort, often utilizing country experts to make the assessments that are then centrally processed. While one criticism of rankings was that their methodology was unclear, the second-generation governance indices are considered transparent in terms of methodological choices (Knack, Kugler, and Manning 2003).

The second-generation governance indicators have provided significant improvements in terms of methodology, providing detailed information on institutions and processes of the countries analyzed, whereas the previous rankings were limited to aggregate figures. The second-generation indicators only provide disaggregated data, which has shifted the focus towards individual institutional aspects, such as the state of transparency.

While the politicization of the ranking technique may have provided the second-generation index producers an opportunity to enter the field, non-aggregate figures are not likely to create news headlines and the information they provide is aimed more at expert audiences. Therefore, the second-generation indicators are not competing for media visibility, which makes it easier to collaborate on data production.

Though the field of governance indices might first appear as a competition among different data producers, the community also shares many normative and causal beliefs on good governance as well as related policy objectives. While the early rankings were informed by the so-called Washington consensus and had institutional ties to the major organizations of economic development, the second-generation indicators are in many ways part of the same movement, now only produced by smaller NGOs. At least in the case of transparency metrics, there are hardly challenges to the ideological premises of the rankings and the second-generation indicators are in many ways reproducing the performance-oriented core beliefs of good governance. In this respect, the critique of ranking may have politicized the aggregate indicators in terms of data presentation, but the attributes of “good governance” as the subject of measurement have remained largely unchallenged.

Moreover, though the shift towards second-generation measurements has provided clear improvements in the methodology of measurements, this has also been limited to the country-level measurements of good governance. As the use of ranking and indicators is spreading to regional and city level, we see similar measurement issues emerging as in the first-generation good governance indicators.

5. MEASUREMENT ISSUES IN REGIONAL AND CITY-LEVEL INDICATORS

As if to counter the above move towards methodologically advanced second-generation indicators, there has been a recent surge of regional and city-level rankings of competitiveness and innovation that are predominantly aggregated composite indicators. Aiming to provide “holistic” measurements, the indicators also contain variables on good governance. Here the problems identified in the early rankings of good governance are again encountered. This can be linked to the field development in global ranking (Erkkilä and Piironen 2018), where knowledge producers are bound by shared normative and causal beliefs, criteria of validity in

Table 11.3 *Regional and city-level rankings on competitiveness and innovation*

	2000–2009	2010–present
Competitiveness	Worldwide Centers of Commerce Index (2007)	EU Regional Competitiveness Index (2010)
	Rich States, Poor States (2007)	Hot Spots 2025 (2013)
	A.T. Kearney’s Global Cities (2008)	The Competitiveness of Cities (2014)
	Global Power City Index (2008)	Global Cities Talent Competitiveness Index (2017)
Innovation	European Innovation Scoreboard (2001)	The Bloomberg Innovation Index (2011)
	Innovation Cities Index (2007)	The Startup Ecosystem Report (2012)
	Innovation Union Scoreboard (2008)	Thomson Reuters Top 100 Global Innovators (2011)
		The Global Cleantech Innovation Index (2012)
		Top 100 Innovative Universities (2015)
		Contributors and Detractors (2016)
	Top 25 Global Innovators – Government (2016)	

assessing the measurements, as well as direct sharing of data. Instead of trying to conceptualize specific aspects of governance and then proceeding to operationalize them in the best possible way, the knowledge producers seek to exploit the existing data sources, at times limiting their own efforts to rearranging the tried data sources into new compositions. Hence the process of knowledge production is deeply characterized by connections between data producers and their sharing of data and methods.

Cities are becoming increasingly important actors in public policy, which explains the rising interest in measuring policies on city and regional level. But it is also important to note that there is now a strong supply side argument driving the chain of development: for the sake of credibility, many international organizations and NGOs are compelled to have a signatory data set of their own (Freistein 2016). The shift to local and regional level also marks an opening for those hoping to enter the ranking activity as a knowledge producer (Erkkilä and Piironen 2018: chap. 6). On the demand side, there are not that many countries or regions desperately wanting to be ranked, though not being ranked at all may also be stigmatic. However, the current field development in global rankings remains largely dominated by the agenda of the small community of index producers, while those measured have very little say on the composition of figures.

Table 11.3 shows selected regional and city-level comparative assessments on competitiveness and innovation. Apart from the European Innovation Scoreboard (published in 2001), all of these measurements have emerged after 2007. In other words, the shift towards the new city-level measurements has occurred around the time that the second-generation indicators emerged. In many ways the city-level indicators could provide an ideal case for fulfilling the criteria of second-generation measurements, as they would allow more focused analysis on a limited area, city, or region. But closer analysis of the measurements reveals that they are all composite indicators drawing heavily from existing figures. In particular, the country comparisons of competitiveness and good governance now provide most of their data. Moreover, the new indicators mostly present their results in an aggregate figure, i.e. they are rankings.

These data sets also serve as a good test case for current methodological problems in indicator knowledge. As composite indicators often developed by small actors such as private companies and consultancies the new data sets are particularly prone to measurement issues. The regional and city-level indicators contain apparent limitations when assessed against the criteria of second-generation indicators, relating to: (1) transparency, (2) availability of data, (3) quality, accuracy, and validity of measurements, and (4) specificity of measurements (Knack, Kugler, and Manning 2003).

Concerning the transparency of measurements, there are problems in documentation of data sources and weighting of data. Some data producers simply do not disclose this information. For instance, the data sources for A.T. Kearney's 2014 Global City Index are shortly summarized as a collection of available data, including other rankings (Leff and Petersen 2015: 12); in the 2015 and 2016 reports, it is merely noted that "sources are derived from publicly available city-level data" (A.T. Kearney 2016: 9). Related to the measurements of transparency presented above, the Global City Index contains a dimension of Information Exchange that constitutes 15% of the rank score (A.T. Kearney 2015). Concerning the indicators used, A.T. Kearney's 2014 data appendix states "information exchange examines how well news and information circulate within and outside the city, based on: accessibility to major television news channels, Internet presence (capturing the robustness of results when searching for the city name in major languages), number of international news bureaus, freedom of expression, and broadband subscriber rate" (A.T. Kearney 2014: 14). While the data source for freedom of expression is not named, there is at least a clear ideational link to the measurements of Freedom House (see above).

In a similar fashion, the governance dimension of A.T. Kearney's Global Cities Outlook contains indicators on transparency, quality of bureaucracy, and ease of doing business, which also bear close ideational resemblance to governance indicators, most notably World Bank's Worldwide Governance Indicators (with quality of governance as one of its dimensions) and Doing Business data set, though the data sources are not named in the report. In its discussion of the 2016 result on Global Cities Outlook, A.T. Kearney provides a link to the Global Innovation Index as further information on global innovation, though this is not named as a data source. And vice versa, the Global Innovation Index 2015 report welcomes A.T. Kearney as its new Knowledge Partner (Cornell University, INSEAD, and WIPO 2015: preface).

There are also actors who regard their data sources and their manipulation and weighting as a proprietary product. For example, the data sources of the Innovation Cities Index by the 2thinknow innovation agency are also not specified and the methodology is a proprietary product that is sold to customers. The Government & Politics segment of the Innovation Cities Index contains many ideational elements similar to the good governance indicators, stressing government responsiveness, transparency, open data and eGovernment initiatives, but again the sources are not named. When discussing its methodology, 2thinknow lists general data types including statistics, commercial sources, rankings, classifications, algorithm-based data, index scores, and estimates. The blunt description of these data types underlines the lack of transparency in data manipulation and weighting data when generating the index. Also, the data sources of Mori Memorial Foundation's Global Power City Index are not available online but are listed in a Yearbook that is being sold for about 120 euros (in 2017).

Linking to the criteria of transparency and non-controversiality, the data used in composite indicators might not be apparent to the end user. For instance, the Global Power City

Index (and many other indicators) contains data from the Washington, DC-based Heritage Foundation that defines itself as a “research and educational institution whose mission is to build and promote conservative public policies”.⁶ The political nature of measurements might not be readily apparent for end users and hence a potential cause for controversy. Though the Heritage Foundation has every right to produce indicators from its own ideological premises, the end users of composite figures on regional and city-level competitiveness should also be made aware of the ideological underpinnings of the used data. It needs to be noted that the organizations with previous experience in producing statistics or indicators tend to document their data sources better. For example, the indicators produced by the European Union are better documented and it is less controversial in its choices of data sources.

There are also apparent problems of data availability. The regional level measurements often build on global-level measurements and use country-level data as a proxy for regional or city level, or make approximations based on it. Though the EU measurements (European Innovation Scoreboard and EU Regional Competitiveness Index) might be of better quality than the rest of the figures, even they are drawing on existing measurements. For example, the Regional Competitiveness Index of the European Union builds on the holistic approach of the WEF’s Global Competitiveness Index (Annoni, Dijkstra, and Gargano 2017: 2). In its assessment of institutions, the EU’s Regional Competitiveness Index uses data from the World Bank’s Worldwide Governance Indicators (all six dimensions), the World Bank’s Ease of Doing Business scores, and eight indicators from the WEF’s Global Competitiveness Index.⁷

Closest to the analysis on “institutions” or “governance”, the “Economy” function of the Global Power City Index (Mori Memorial Foundation) also draws sources from existing indicators such as the World Bank’s Ease of Doing Business, Moody’s credit rating, the Heritage Foundation’s Index of Economic Freedom, and the Global Talent Competitiveness Index by INSEAD. In particular, the WEF’s Executive Opinion Survey (underlying the Global Competitiveness Index) and the World Bank’s Doing Business data sets are broadly used in the regional and city-level measurements. Also country-level data from UN organizations and the OECD are commonly used. This is problematic, as the city-level measurements in fact contain a very limited amount of actual city-level information but rather depend heavily on country-level data.

The problem becomes more pressing with data producers such as A.T. Kearney and 2thinknow who do not reveal their data sources. A.T. Kearney currently simply disclaims that “in the few cases when city-level data is not available, country-level data is used” (A.T. Kearney 2016: 9), while 2thinknow notes that “[w]ithout 2thinknow process of standardization of rankings for City Benchmarking Data, data may not be comparable across cities. We also take national rankings and through our standard process acclimatise and adapt them to city indicators.”⁸ The company states that information on data sources and methodology is proprietary and only available to paying customers.⁹

In short, the city-level measurements draw heavily on national-level data and occasionally fail to report this in the public documentation of their measurements. Many indicator producers acknowledge this as a potential limitation of their measurement. But at the other extreme, 2thinknow perceives the manipulation and compiling of various types of data into city-level information as its commercial product. The continuity over time might also be a problem as many of the data sources are new. The data producers of city rankings have also been reported to have used data from different years (Leff and Petersen 2015). The above examples show

how the new city-level measurements are struggling to get data on cities and are compelled to use country data instead, or at least to make estimates based on it.

The relatively youthful nature of city rankings becomes apparent when we look at how the above categories have been conceptualized and what data and methodology have been used. There is a certain resemblance to the early measurements of good governance, which were composite indicators, using data from various sources. In the absence of data, the city-level measurements mostly use available public data sources, instead of producing data themselves, leading them to use country data in the absence of city-level data (Leff and Petersen 2015).

Also the quality and accuracy of data has great variation. The EU measurements are by far the most qualitative ones, based mostly on Eurostat data. The other producers of regional and city-level indicators obviously lack similar data sources. Furthermore, the new regional and city-level measurements are not very specific in nature, as they often aim at making general assessments of very abstract concepts of innovation and competitiveness. The shift towards second-generation indicators that we witness in the country rankings finds no correspondence on the city level. While the global indicators are improving in terms of accuracy and quality of data, most of the measurements that have emerged on the regional and city level are suffering from the limitations identified previously in the country rankings.

The emerging sub-national indicators also lack specificity – it is quite unclear what is being measured. This owes to the poor conceptualization of broad categories such as competitiveness, innovation, start-up ecosystems, or “power city”. In principle, the closer geographic focus would allow the city rankings to adopt the principles of second-generation indicators aiming for specificity. But the result is often the opposite. Many of the measurements are poorly conceptualized and based on publicly available data of various types. The producers have limited resources to produce data themselves. Moreover, the criteria for second-generation indicators such as transparency of data sources and weighting of indicators might even conflict with the commercial interests of those producing the city-level data.

Second-generation indicators are also often referred to as “mappings”, as they allow different (disaggregated) representations of data, instead of just a single aggregate number. So far, the methodological critique against the city-level measurements has not been as explicit as in the context of country rankings. But the decision to create rankings has been debated to some extent. While the global-level indicators of good governance are moving towards non-aggregated measurements, the city-level measurements produced at about the same time are mostly aggregated rankings. Interestingly, in its *Competitiveness of Cities* report, the World Economic Forum distances itself from previous indicators on city competitiveness, promoting case studies (World Economic Forum 2014: 7).

In 2015, the World Bank published a report on competitive cities, arguing that there is a growing consensus about their global importance (World Bank 2015: 19). The report reviews available measurements of city competitiveness and argues that the existing indicators have a rich country bias, pointing to potential political controversy. The report further suggests disaggregation of the indexes and using new data to supplement the existing measures (World Bank 2015: 37). This shows how the ideas of disaggregation and second-generation indicators are now promoted by actors that themselves have been criticized for the very same limitations earlier. To be fair, the World Bank has also shown reflexivity over its country-level measurements, launching its Actionable Governance Indicators, a mapping instrument (see above).

While this could simply be an attempt to draw attention to their comparative work on the city level, this debate could provide an opening for disaggregated indicators comparing cities.

Furthermore, this could also allow different conceptions of a competitive city and taking better into account the concerns of the global South (World Bank 2015: 23). Particularly the BRICS countries are keenly following the existing figures and tend to see them poorly fitting their institutional conditions, arguing for their adjustment (Pulakkat 2016; *The Economic Times* 2016; World Economic Forum 2017). This has led to initiatives such as the India Innovation Index that arguably would improve the existing “data gaps” by merging the Global Innovation Index with India-specific data (*The Economic Times* 2017).

There is also reflexivity about the figures among their producers. The Global Cities Talent Competitiveness Index produced by INSEAD business school with the Adecco Group and the Human Capital Leadership Institute of Singapore was in its first edition (2017) still based on the country-level measurement Global Talent Competitiveness Index by the same producers. The next iteration of this city rankings in 2018 contained significantly fewer indicators and also acknowledged problems in the initial data sets, leading to the “streamlining” of the model and introduction of an updated set of variables, resulting from “abundant comments and feedback” (INSEAD 2018: 89). The 2018 iteration also discusses the problems of data availability and applicability that arguably were taken into account when deciding on the composition of the indicator (INSEAD 2018: 90).

The above examples point to the need of including a feedback loop in the development of data sets, including potentially also the subjects of measurement. What sets apart the new types of indicator and ranking knowledge from the earlier international statistics is the heterogeneous nature of their producers. While international statistics were previously often created by international organizations such as the UN, World Bank, OECD, and the EU that were closely collaborating with nation states and their statistical bureaus, the new governance indicators particularly on the city level are mostly produced by organizations that have limited contact with the subjects of their measurement, let alone means of representation or venues of deliberation concerning the measurements.

6. SUMMARY

The above examples highlight the need for agreed upon criteria for producing indicator knowledge. The criteria for second-generation indicators that have been used in this chapter provide a starting point. But there is still a long way to go in terms of commitment to upholding these. Similar attempts have been made in the field of university rankings. Promoted by major producers of university rankings, the so-called Berlin principles¹⁰ have informed the statistical work in this policy domain since 2006, though these have also been criticized (Hägg and Wedlin 2013).

The most important step in the process of improving the indicators would be to aim for precise and focused concepts that apply to different contexts (Sartori 1970). In addition, one needs to be self-conscious and reflexive about the conceptual innovation in comparative research (Collier and Levitsky 1997). At present, the comparative analysis of index producers is more focused on the availability of data than on concepts that “travel” (Sartori 1970). The new addendum of city-level analysis is mainly conceived with the help of existing indicators that are grouped into composite measurements that often stretch country-level data to make arguments about urban areas.

7. CONCLUSIONS

In this chapter I have analyzed measurement issues in global indicators, discussing the rise of second-generation governance indicators and related methodological debates and changes in the production and use of indicators. As the case of transparency measurements shows, there has been a shift towards more detailed and methodologically improved data sets that comply with the criteria set for the second-generation indicators. However, more recently, there has been a surge of regional and city-level rankings of competitiveness and innovation that aim for “holistic” measurements through composite indicators and aggregation. Here the problems identified in the early rankings of good governance are again encountered. Critical appraisal of regional and city-level indicators shows clear limitations in the criteria of second-generation indicators.

These measurement issues can be understood against the field development in global ranking, where the actors entering the field often come not only to share the prevailing epistemic practices, but also the data readily available. While international statistics was previously a domain reserved mainly for international organizations, often with the backing of national statistical bureaus, we now see private companies and international NGOs engaging comparative numerical assessments of global scope with limited resources while facing pressure to establish expert status through a signatory data set. Consequently, the global comparative assessments do not travel on their concepts (Sartori 1970), but on availability of existing data. This now leads to conceptual stretching, but is also related to various other measurements issues that prevail despite attempts at establishing criteria for second-generation indicators.

NOTES

1. These new types of indicators are also called “actionable” governance indicators because they allow close monitoring and development of specific aspects of governance (Trapnell 2011; World Bank 2009).
2. Government transparency is generally understood as access to government documents, involving also public information on statistics, budget, and government performance, as well as an uncensored internet, reflected in the indicators selected for analysis.
3. The WGI do not provide a single ranking but instead only results for six aggregate indicators.
4. Nevertheless, as with WGI, it measures governance in terms of its economic qualities (Erkkilä and Piironen 2014) and the assessments of transparency are centered on the notion of budget transparency, understood as the collection, allocation, and use of performance information.
5. At present, there seems to be a grassroots-level development in the Washington, DC area, where many NGOs now work on transparency issues.
6. Heritage Foundation: <https://www.heritage.org/>. (accessed June 2018).
7. EU Regional Competitiveness Index 2016: indicators description. http://ec.europa.eu/regional_policy/sources/docgener/work/rci2016_indicators.xls. (accessed June 2018).
8. 2thinknow Data Types: <http://www.citybenchmarkingdata.com/data-types>; 2thinknow FAQs: <http://www.citybenchmarkingdata.com/faqs/2905>. (accessed June 2018).
9. 2thinknow FAQs: <http://www.citybenchmarkingdata.com/faqs/2905>. (accessed June 2018).
10. CHE Berlin Principles: https://www.che.de/downloads/Berlin_Principles_IREG_534.pdf. (accessed June 2018).

REFERENCES

- Andrews, M. (2008). The good governance agenda: beyond indicators without theory. *Oxford Development Studies* 36 (4): 379–407.
- Andrews, M., Hay, R., and Myers, J. (2010). Can governance indicators make sense? Towards a new approach to sector-specific measures of governance. *Oxford Development Studies* 38 (4): 391–410.
- Annoni, P., Dijkstra, L., and Gargano, N. (2017). *The EU Regional Competitiveness Index 2016*. Working Paper 02/2017. European Commission. http://ec.europa.eu/regional_policy/sources/docgener/work/201701_regional_competitiveness2016.pdf. (accessed June 2018).
- A.T. Kearney (2014). *Global Cities 2014*. Chicago, IL: A.T. Kearney.
- A.T. Kearney (2015). *Global Cities 2015*. Chicago, IL: A.T. Kearney.
- A.T. Kearney (2016). *Global Cities 2016*. Chicago, IL: A.T. Kearney.
- Barendrecht, M. (2011). Rule of law, measuring and accountability: problems to be solved bottom up. *Hague Journal on the Rule of Law* 3 (2): 281–304.
- Buduru, B. and Pal, L. A. (2010). The globalized state: measuring and monitoring governance. *European Journal of Cultural Studies* 13 (4): 511–30.
- Collier, D. and Levitsky, S. (1997). Democracy with adjectives: conceptual innovation in comparative research. *World Politics* 49 (3): 430–451.
- Cornell University, INSEAD, and WIPO (2015). *The Global Innovation Index 2015: Effective Innovation Policies for Development*.
- Davis, K. E., Kingsbury, B., and Merry, S. E. (2012). Indicators as a technology of global governance. *Law & Society Review* 46: 71–104.
- Desrosières, A. (1998). *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge, MA: Harvard University Press.
- Drechsler, W. (2004). Governance, good governance, and government: the case for Estonian administrative capacity. *TRAMES* 4: 388–96.
- Erkkilä, T. (Ed.) (2013). *Global University Rankings: Challenges for European Higher Education*. Basingstoke: Palgrave Macmillan.
- Erkkilä, T. (2016). Global governance indices as policy instruments: actionability, transparency and comparative policy analysis. *Journal of Comparative Policy Analysis: Research and Practice* 18 (4): 382–402.
- Erkkilä, T. and Piironen, O. (2009). Politics and numbers: the iron cage of governance indices. In R. W. Cox III (Ed.), *Ethics and Integrity of Public Administration: Concepts and Cases*. Armonk, NY: M. E. Sharpe, 125–45.
- Erkkilä, T. and Piironen, O. (2014). (De)politicizing good governance: the World Bank Institute, the OECD and the politics of governance indicators. *Innovation: The European Journal of Social Science Research* 27 (4): 344–60.
- Erkkilä, T. and Piironen, O. (2018). *Rankings and Global Knowledge Governance: Higher Education, Innovation and Competitiveness*. New York: Palgrave Macmillan.
- Espeland, W. N. and Sauder, M. (2007). Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology* 113 (1): 1–40.
- Freistein, K. (2016). Effects of indicator use: a comparison of poverty measuring instruments at the World Bank. *Journal of Comparative Policy Analysis: Research and Practice* 18 (4): 366–81.
- Ginsburg, T. (2011). Pitfalls of measuring the rule of law. *Hague Journal on the Rule of Law* 3 (2): 269–80.
- Global Integrity (2011). *The Global Integrity Report: 2011. Methodology White Paper*. https://www.globalintegrity.org/wp-content/uploads/2013/08/2011_GIR_Meth_Whitepaper.pdf. (accessed June 2018).
- Gornitzka, Å. (2013). Channel, filter or buffer? National policy responses to global rankings. In T. Erkkilä (Ed.), *Global University Rankings: Challenges for European Higher Education*. Basingstoke: Palgrave Macmillan, 75–91.
- Gramatikov, M., Barendrecht, M., and Verdonschot, J. H. (2011). Measuring the costs and quality of paths to justice: contours of a methodology. *Hague Journal on the Rule of Law* 3 (2): 349–79.

- Guha, K. and McGregor, R. (2007). World Bank directors test Zoellick. *Financial Times*. <http://www.ft.com/intl/cms/s/0/fe1d7ece-30d8-11dc-9a81-0000779fd2ac.html#axzz2883nwxuM>. (accessed June 2018).
- Haas, P. M. (1992). Introduction: epistemic communities and international policy coordination. *International Organization* 46 (1): 1–35.
- Hägg, I. and Wedlin, L. (2013). Standards for quality? A critical appraisal of the Berlin principles for international rankings of universities. *Quality in Higher Education* 19 (3): 326–42.
- Harding, R. (2013). China seeks to water down key World Bank report. *Financial Times*. <http://www.ft.com/intl/cms/s/0/a1ca36d2-b654-11e2-b1e5-00144feabdc0.html#axzz2SmcKhrYs>. (accessed June 2018).
- Hazell, R. and Worthy, B. (2010). Assessing the performance of freedom of information. *Government Information Quarterly* 27 (4): 352–9.
- Hinthorne, L. L. (2011). Democratic crisis or crisis of confidence? What local perceptual lenses tell us about Madagascar's 2009 political crisis. *Democratization* 18 (2): 535–61.
- Hood, C. and Margetts, H. (2007). *The Tools of Government in the Digital Age*. Basingstoke: Palgrave Macmillan.
- INSEAD (2018). *The Global Talent Competitiveness Index 2018: Diversity for Competitiveness*. <https://gtcistudy.com/wp-content/uploads/2018/01/GTCI-2018-web.r1-1.pdf>. (accessed June 2018).
- Ivanov, K. (2009). Fighting corruption globally and locally. In R. W. Cox III (Ed.), *Ethics and Integrity in Public Administration: Concepts and Cases*. Armonk, NY: M. E. Sharpe, 146–54.
- Joshi, D. (2011). Good governance, state capacity, and the Millennium Development Goals. *Perspectives on Global Development & Technology* 10 (2): 339–60.
- Kaufmann, D., Kraay, A., and Mastruzzi, M. (2008). *Governance Matters VII: Aggregate and Individual Governance Indicators, 1996–2007*. World Bank Policy Research Working Paper No.4654. Washington, DC: World Bank.
- Kaufmann, D., Kraay, A., and Mastruzzi, M. (2010). Response to “What do the worldwide governance indicators measure?” *European Journal of Development Research* 22 (1): 55–8.
- Kaufmann, D., Kraay, A., and Mastruzzi, M. (2011). The worldwide governance indicators: methodology and analytical issues. *Hague Journal on the Rule of Law* 3 (2): 220–246.
- Kauppi, N. and Erkkilä, T. (2011). The struggle over global higher education: actors, institutions, and practices. *International Political Sociology* 5 (3): 314–26.
- Kelley, J. G. and Simmons, B. A. (2015). Politics by number: indicators as social pressure in international relations. *American Journal of Political Science* 59 (1): 55–70.
- Knack, S., Kugler, M., and Manning, N. (2003). Second-generation governance indicators. *International Review of Administrative Sciences* 69 (3): 345–64.
- Langbein, L. and Knack, S. (2010). The worldwide governance indicators: six, one, or none? *Journal of Development Studies* 46 (2): 350–370.
- Leff, S. and Petersen, B. (2015). *Beyond the Scorecard: Understanding Global City Rankings*. Chicago, IL: Chicago Council on Global Affairs.
- Löwenheim, O. (2008). Examining the state: a Foucauldian perspective on international “governance indicators”. *Third World Quarterly* 29 (2): 255–74.
- McFerson, H. M. (2009). Measuring African governance. *Journal of Developing Societies* 25 (2): 253–74.
- Merry, S. E. (2011). Measuring the world. *Current Anthropology* 52 (3): S83–S95.
- Merry, S. E., Davis, K. E., and Kingsbury, B. (Eds.) (2015). *The Quiet Power of Indicators*. New York: Cambridge University Press.
- Michener, G. (2015). Policy evaluation via composite indexes: qualitative lessons from international transparency policy indexes. *World Development* 74: 184–96.
- Miller, P. and Rose, N. (1990). Political rationalities and technologies of government. In S. Hänninen and K. Palonen (Eds.), *Texts, Contexts, Concepts: Studies on Politics and Power in Language*. Helsinki: The Finnish Political Science Association, 166–83.
- Morgan, G. (2011). Traction on the ground: from better data to better policy. *Hague Journal on the Rule of Law* 3 (2): 380–396.
- Neumann, R. and Graeff, P. (2010). A multitrait-multimethod approach to pinpoint the validity of aggregated governance indicators. *Quality & Quantity* 44 (5): 849–64.

- OECD (2006). *How and Why Should Government Activity Be Measured in "Government at a Glance"?* OECD GOV Technical Paper 1. GOV/PGC(2006)10/ANN1. Paris: OECD.
- OECD (2007). *Towards Better Measurement of Government*. OECD Working Papers on Public Governance, 2007/1. Paris: OECD.
- Pidd, M. (2005). Perversity in public service performance measurement. *International Journal of Productivity and Performance Management* 54 (5–6): 482–93.
- Porter, T. M. (1996). *Trust in Numbers*. Princeton, NJ: Princeton University Press.
- Pulakkat, H. (2016). A look at India's 15-place jump in global innovation index. *The Economic Times*. <http://economictimes.indiatimes.com/news/economy/indicators/a-look-at-indias-15-place-jump-in-global-innovation-index/articleshow/53852329.cms>. (accessed June 2018).
- Robinson, P. (2003). Government accountability and performance measurement. *Critical Perspectives on Accounting* 14 (1): 171–86.
- Sartori, G. (1970). Concept misformation in comparative politics. *American Political Science Review* 64 (4): 1033–53.
- Seifert, J., Carlitz, R., and Mondo, E. (2013). The Open Budget Index (OBI) as a comparative statistical tool. *Journal of Comparative Policy Analysis: Research and Practice* 15 (1): 87–101.
- Seppänen, S. (2003). *Good Governance in International Law*. Helsinki: The Erik Castrén Institute.
- Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration* 18 (2–3): 277–310.
- Stiglitz, J. (1998). Distinguished Lecture on Economics in Government: The private uses of public interests: incentives and institutions. *Journal of Economic Perspectives* 12 (2): 3–22.
- Stiglitz, J. (2002). Information and the change in the paradigm in economics. *American Economic Review* 92 (3): 460–501.
- The Economic Times (2016). Modify global innovation index methodology: India. *The Economic Times*. <http://economictimes.indiatimes.com/news/politics-and-nation/modify-global-innovation-index-methodology-india/articleshow/53776420.cms>. (accessed June 2018).
- The Economic Times (2017). Niti Ayog launches "India innovation index". *The Economic Times*. <http://economictimes.indiatimes.com/news/economy/policy/niti-ayog-launches-india-innovation-index/articleshow/56940608.cms>. (accessed June 2018).
- Thiel, S. van and Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance & Management Review* 25 (3): 267–81.
- Thomas, M. A. (2010). What do the worldwide governance indicators measure? *European Journal of Development Research* 22 (1): 31–54.
- Trapnell, S. E. (2011). Actionable governance indicators: turning measurement into reform. *Hague Journal of the Rule of Law* 3 (2): 317–48.
- van de Walle, S. (2006). The state of the world's bureaucracies. *Journal of Comparative Policy Analysis: Research and Practice* 8 (4): 437–48.
- World Bank (2007). *Strengthening World Bank Group Engagement on Governance and Anticorruption*. Washington, DC: World Bank.
- World Bank (2009). Actionable governance indicators | data. <http://data.worldbank.org/data-catalog/actionable-governance-indicators>. (accessed June 2018).
- World Bank (2015). Competitive cities for jobs and growth: what, who and how. http://www-wds.worldbank.org/external/default/WDSPContentServer/WDSP/IB/2016/04/21/090224b0842b4f79/3_0/Rendered/PDF/Competitive0ci000what00who00and0how.pdf. (accessed June 2018).
- World Economic Forum (2014). *The Competitiveness of Cities: A Report of the Global Agenda Council on Competitiveness*. Geneva: World Economic Forum.
- World Economic Forum (2017). New index to measure innovation performance of Indian states. <https://www.weforum.org/press/2017/01/new-index-to-measure-innovation-performance-of-indian-states/>. (accessed June 2018).
- Zanotti, L. (2005). Governmentalizing the post-Cold War international regime: the UN debate on democratization and good governance. *Alternatives* 30 (4): 461–87.

12. Using text-as-data methods in comparative policy analysis

Fabrizio Gilardi and Bruno Wüest

1. INTRODUCTION

Text-as-data methods are a broad set of techniques and approaches relying on the automated or semi-automated analysis of text. They have become increasingly prevalent in the social sciences, and are part of a broader trend in which, taken together, the internet and computational social science tools have changed the kinds of questions that social scientists can ask and answer successfully (Golder and Macy 2014; Lazer and Radford 2017). Text analysis holds a prominent place in these developments. Texts have always been a primary data source for social scientists. As Monroe and Schrodt (2008: 351) write, “text is arguably the most pervasive – and certainly the most persistent – artifact of political behavior”. In the internet age, texts have become particularly plentiful, and accessible with relative ease. The large amount of text available to researchers, combined with new computational tools, has promoted the development of text-as-data approaches in which texts are analyzed statistically with different degrees of automatization. The promise of the approach is that it can both apply existing theories to new data and uncover new phenomena that previously remained hidden (Evans and Aceves 2016). As González-Bailón (2017: xviii) writes: “when the right connections are made, much of the data-driven research that is being conducted today speaks directly to long-standing (and unresolved) theoretical discussions”.

Text-as-data approaches are becoming mainstream in political science. Typical applications revolve around research questions where at least one element is based on political communication theories such as agenda setting, issue definition, or framing (for reviews, see Grimmer and Stewart 2013; Lucas et al. 2015; Wilkerson and Casas 2017). From a practical perspective, these approaches allow researchers to conduct more efficiently research they have been doing manually for decades, such as classifying texts into categories such as policy areas. But, thanks to powerful inductive analysis procedures, text-as-data approaches also make it possible to discover new phenomena, concepts, and relationships from latent dimensions of texts.

The focus of this chapter is on comparative policy analysis more specifically, a field in which text-as-data methods have not yet been applied widely despite the potential demonstrated by applications in other subfields. Textual materials covering various administrative, legislative, and political aspects of policy processes have always been a central data source for policy analysis. For example, texts have been the central source for the successful and influential Comparative Agendas Project (Baumgartner, Green-Pedersen, and Jones 2006; Baumgartner, Jones, and Wilkerson 2011; Dowding, Hindmoor, and Martin 2016; John 2006), which originally relied on manual coding to classify legislation and other relevant texts into 21 major topics and 220 subtopics,¹ but is increasingly using automated approaches to carry out this task.

The specificity of text-as-data-methods for comparative policy analysis lies in the application of existing tools and approaches to both classic and new theories and phenomena relevant for public policy and policy analysis. However, it is possible, and certainly desirable, that scholars of policy analysis will build on existing methods to adapt them to their specific research needs.

The goal of the chapter is to offer an overview of existing applications and, especially, of the options and workflow of text-as-data approaches for comparative policy analysis. The learning curve can at first appear quite steep for these methods, and we hope to motivate policy analysts to take on the challenge by showing the potential payoffs as well as offering an overview of the various practical steps and available options. To this end, the chapter first reviews applications of text-as-data methods in comparative policy analysis. Then, it focuses on the practical aspects of these methods, and specifically on the workflow involved in their application, such as obtaining and storing the data, preprocessing, and analyzing them with a range of automated and semi-automated techniques. Next, it presents three specific kinds of applications: concept identification, classification, and discovery. The chapter concludes by highlighting the potential of text-as-data methods for comparative policy analysis despite their relatively sparse use so far.

2. TEXT-AS-DATA APPLICATIONS IN COMPARATIVE POLICY ANALYSIS: AN OVERVIEW

The use of text-as-data techniques in political science has increased steadily in the past years and has become highly diversified. The origins of text-as-data methods can be tracked to various approaches such as classical content analysis (Krippendorf 2004) and the computer science literature on natural language processing (Jurafsky and Martin 2009). In political science, event detection systems for conflict studies are among the earliest applications (Gerner et al. 1994). Originally, these approaches used keyword matching to encode events such as bomb attacks. Later, these approaches were enhanced with more advanced tools such as syntax parsing, lexical databases, and named entity recognition tools. Other early applications in political science relied on scaling approaches, which involves a broad array of methods aiming to map texts on one or more underlying dimensions. Starting with the Wordscore method (Laver, Benoit, and Garry 2003; Lowe 2008) and the widely used Wordfish method (Slapin and Proksch 2008), these approaches have been used primarily to extract ideological or policy-specific ideal points from texts such as legislative speeches and party manifestos.

In the last decade, a rapidly growing literature has aimed to develop tools for classifying political texts (e.g. Hopkins and King 2010). Most commonly, politically meaningful texts are classified into issue or topic categories. Such classifications can be conducted deductively using machine learning, or they can be conducted inductively with latent variable models. Examples include studies that analyze policy-specific debates in parliaments (Quinn et al. 2010), censorship by authoritarian regimes (King, Pan, and Roberts 2013), electoral representation (Grimmer 2013), and individual opinion and decision making (Wüest 2018). The rise of social media platforms such as Facebook and Twitter for political communication has further opened the way for social network and big data analysis into political science (for an overview, see Jungherr and Theocharis 2017).

In comparative policy analysis, applications of text-as-data methods are relatively rare, despite the potential demonstrated by their increasing popularity in other related fields. To be sure, qualitative text analysis is prevalent in the literature, sometimes in combination with quantitative methods. For example, the discourse networks approach relies on text analysis to measure discourse coalitions quantitatively through network analysis (Fisher, Leifeld, and Iwaki 2013; Fisher, Waggle, and Leifeld 2013; Leifeld 2013; Leifeld and Haunss 2012). Although text-as-data applications are not yet mainstream in comparative policy analysis, a few notable exceptions have addressed questions that have been at the core of policy analysis for many years.

2.1 Policy Agendas

The Comparative Agendas Project is a network of researchers developing a measurement system to classify a broad range of political activities into topics which can be compared over time and across political systems (Baumgartner, Breunig and Grossman 2019). The project builds on Baumgartner and Jones' pioneering work on policy agendas (Baumgartner and Jones 1993; Jones and Baumgartner 2005). Initially, and for many years, the project employed manual coding to classify a wide range of legislative, judiciary, and journalistic texts into policy categories, using a very detailed coding scheme. More recently, the project has started to rely on machine-learning procedures in which manually prepared training data sets are fed to classification algorithms (Collingwood and Wilkerson 2012). The move to automatized procedures was needed particularly in countries or areas that were not part of the original project or for which the costs of manual coding were prohibitive, such as Hungary (Sebók and Berki 2017) and Danish city councils (Loftis and Mortensen 2018). These examples illustrate one of the main advantages of text-as-data methods, namely, the possibility to extend existing projects to new areas at a relatively low cost.

2.2 Problem Definition

A textbook argument in policy analysis is that the decision-making process consists of several stages, and that the stage in which problems are defined as politically relevant is a crucial one. As Elder and Cobb (1984: 115) noted, because “policy problems are not a priori givens but rather are matters of definition ... what is at issue in the agenda-building process is not just which problems will be considered but how those problems will be defined”. The idea that problem definition affects the kinds of policies that are adopted, as well as those that are not, is now considered “nearly axiomatic” within the policy making literature (Boushey 2016: 200). Texts are a natural source to study problem definition, and some studies have started to rely on text-as-data approaches to do so. Nowlin (2016) shows how topic models can be used to study how issues are defined and applies the approach to Congressional hearings regarding used nuclear fuel. His analysis identifies seven dimensions (programmatic, safety/regulation, Yucca mountain, site selection, science/technical, storage, and transportation) and shows that their salience co-varies with important policy developments. Gilardi, Shipan, and Wüest (2020) apply topic models to newspaper articles on smoking bans in US states and find significant differences in how the issue was defined, both across states and over time. They employ structural topic models to extract the issue definitions and, at the same time, to estimate the correlation of the issues with co-variates such as the sentiment with respect to smoking bans.

The Policy Frames Project, finally, uses machine learning to track media tone and framing in a variety of areas (Card et al. 2015, 2016), with the aim of providing a comprehensive scheme for the identification of policy frames. This example illustrates another advantage of the text-as-data approach, namely, the possibility to build on other researchers' work at sharply decreasing marginal costs.

2.3 Policy Diffusion

Policy diffusion refers to the phenomenon whereby policies in one unit (city, state, country, etc.) are influenced by policies in other units (Braun and Gilardi 2006; Graham, Shipan, and Volden 2013; Simmons, Dobbin, and Garrett 2006). It is a classic question going back at least to Walker (1969), which in the policy analysis literature is also studied under the label "policy transfer" (Dolowitz 2000; Dolowitz and Marsh 1996). Traditionally, the focus has been on policy adoptions, but scholars have increasingly been interested in how diffusion affects other aspects of the policy making process. For example, Gilardi, Shipan, and Wüest (2020) study how issue definitions diffuse across US states, finding that practical aspects of smoking restrictions are more subject to diffusion than normative rationales. Wilkerson, Smith, and Stramp (2015) use a text reuse approach to trace how ideas spread from one piece of legislation to another, as well as from initial drafts to final bills, which helps uncover the influence of specific lawmakers. Linder et al. (2018) show how the approach can be used to measure policy similarity more in general, especially in a diffusion context. These works illustrate how text-as-data methods can be used to study classic questions from new angles that were previously impractical due to technical constraints.

2.4 Lobbying

Klüver (2013) applies Wordfish (Slapin and Proksch 2008), a scaling method to reduce the dimensionality of texts, to measure the policy preferences of interest groups based on their submissions in online consultation of the European Commission. The analysis underscores the collective nature of lobbying, where success depends on the interaction between information supply, citizen support, and economic power of lobbying camps. Again using consultations, Klüver and Mahoney (2015) and Klüver, Mahoney, and Opper (2015) use cluster analysis to identify the frames used by interest groups as well as their determinants and effectiveness. Klüver, Mahoney, and Opper (2015) show how interest groups tailor their frames based on their targets, while Klüver and Mahoney (2015) show how the European Commission has adopted the frames put forward by various lobbies. This stream of research shows how text-as-data methods have been integrated in a well-established literature.

2.5 Policy Feedbacks

The feedback effects of policies, and specifically how policies affect political dynamics, is a classic question (Pierson 1993) that is crucial to path dependence theories (Pierson 2000). Flores (2017) studies this question using a dictionary approach to measure the sentiment of tweets to identify how public opinion reacts to anti-immigrant legislation. The study finds that the policy affected public discourse not by changing attitudes, but by mobilizing people

already critical of immigrants. This is again an example of how text-as-data methods can shed new light on long-standing discussions in policy analysis.

2.6 Summary

Most of these approaches have in common that they rely on statistical algorithms that estimate the quantities of interests from a bag-of-words representation – basically word frequency distributions – of the texts. This also implies an application of basic concepts of statistical learning such as training and cross-validation. In addition, most of the text-as-data applications discussed above use large or even very large text collections that cannot be analyzed manually with reasonable effort. Another characteristic of these studies is that they rely on existing approaches, which they apply to a specific research question. In other words, text-as-data applications in comparative policy analysis take advantage of methods developed in other subfields. Policy analyses adjust the methods to their specific needs, but they have not created new text-as-data approaches as such.

3. TEXT-AS-DATA IN PRACTICE

Text-as-data approaches involve many moving parts and can be highly complex. On the side of prediction, applications of automated text analysis rely primarily on statistical or heuristic algorithms that retrieve information from bag-of-words representations of the original texts. That is, many kinds of information conveyed in text documents, from morphological information such as word order and word ambiguities to semantic information such as irony, metaphors, or double negatives, are often discarded in the preprocessing steps prior to the analysis. Although such simplifications can appear unreasonable, they have been shown to be surprisingly effective for many kinds of applications. However, they also imply that all results of automated models of language at best are useful approximations of the quantities of interest (Grimmer and Stewart 2013). Moreover, they imply that automated models of language necessarily are highly domain-specific. Accordingly, there are no globally best methods for retrieving certain information from texts. A careful selection and validation of both the preparatory steps and the methods of analysis are therefore essential for the success of a research project. Whenever possible, the validity of text-as-data applications should always be quantified and reported using commonly accepted measures such as *recall*, *precision*, and Cohen's κ .

Figure 12.1 shows the main steps of a text-as-data application, which will be described in more detail in the following sections. A first aspect that is often overlooked are the procedures to obtain the data. Then, researchers need to decide how the documents should be preprocessed. Finally, different research goals imply different methodological strategies, so researchers need to make the theoretically and practically appropriate choices in terms of the methods applied.

The following discussion will first present the preparatory steps, before the various methods are outlined in section 4.

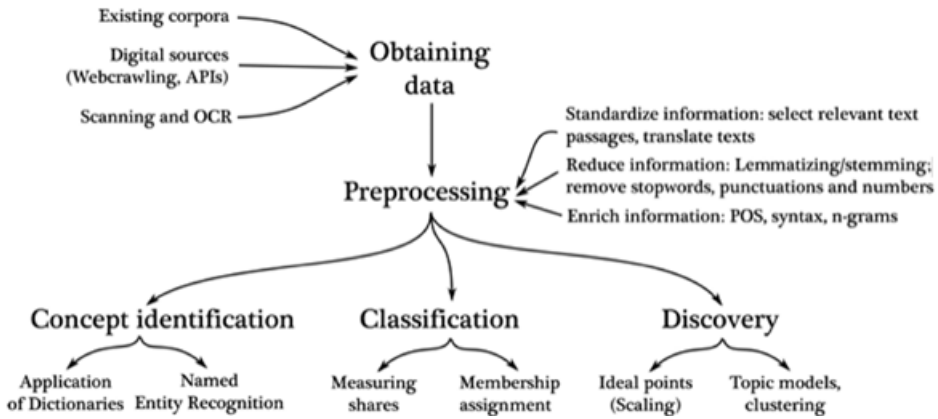


Figure 12.1 Overview of text-as-data applications

3.1 Preparing a Text Corpus: Obtaining the Data

The unheralded first step in text-as-data projects, the construction of a text corpus, usually requires a lot of effort. This is because many documents of interest for researchers in public policy are not easily accessible. For instance, it may come as a surprise that about a third of the state legislatures in the United States do not regularly publish their floor debates on the internet. In some cases, it is simply a question of accessibility, and can be resolved by contacting the database provider. In most cases, however, there do not even exist digital versions of the documents of interest. More often than not, historical archives have not been digitized, and it is often these archives that are of particular importance to carry out diachronic policy analysis. In this scenario, the only option is to climb into the archives, to scan the documents and run an Optical Character Recognition (OCR) software on the scans. Depending on the quality of the documents – e.g. the fonts’ sharpness and the contrast between fonts and sheets – this step can already require a lot of effort.

In addition, the terms and conditions of many database providers are often less than conducive to large-scale text mining projects. First, the usual web interfaces mostly do not allow bulk downloads of texts, either because it is explicitly prohibited or because web scraping is too slow to retrieve a large number of documents within a reasonable time frame. In other words, if the search and download in a web interface takes several seconds and only a few texts can be downloaded at once, it easily takes months until a large corpus is compiled. Possible solutions are programmable interfaces (API) to the providers’ database or a special agreement for a one-time transfer of large data can be negotiated. Second, text mining on the retrieved documents often is prohibited. Most newspapers, for example, explicitly prohibit automated information extraction from their articles in their terms and conditions. In addition, most original data from commercial database providers cannot be published. Obviously, this runs contrary to both the scientific principle of reproducibility and the open data policy of many publishers. In practice, this is an extensive gray area, with one position arguing that almost everything can and should be published anyway – e.g. that documents with one word removed from the text do not count as original data anymore – and another position holding that explicit permission needs to be granted by the database provider (Tennant et al. 2016). To mitigate any

uncertainties, approaching the database providers and transparently negotiating the terms of analysis and publication seems the most promising way. Here, university bodies such as the central library can provide essential support.

As for the storing of the data, we would recommend a different infrastructure depending on the capabilities and resources available to the researcher. Optimally, a large text corpus should be stored in non-relational database such as *ElasticNet* or *MongoDB*, which allow for efficient storage as well as fast document searches (Jatana et al. 2012). This set-up, however, needs particular IT skills and a server infrastructure that may not be available in every research team in comparative policy analysis. An alternative, low-threshold solution is the storage of a corpus in single text files that are systematically stored in a folder tree (e.g. organized by source and date of publication). This means longer times to load the corpus, but it may be easier to keep an overview of the data for researchers not used to working with large text data.

3.2 Preprocessing

Given the inherently unstructured nature of text data, it needs careful preparation before it can be analyzed. The preprocessing of text data involves three steps and can be carried out using several tools such as *spacyR*, *TM*, *quanteda* or *udpipe* in R, or *NLTK*, *spacy* and *polyglot* in Python.

First, when preparing a corpus, researchers need to invest in the standardization of the texts. Several technical details such as the standardization of character encodings and the extraction of meta-data (publication dates, authors, etc.) need to be clarified. Especially if the documents stem from different sources, the encoding may vary depending on the operating systems and software programs used to process these documents. We recommend standardizing texts into one of the most common encodings suitable also for special characters such as German umlauts, such as *utf8* or *latin1*. Meta-data, on the other hand, include information, such as the source or the author of the documents, that can be very important for the analysis. Standardization also involves the definition of the units of observation, that is, the relevant text passages. For example, if parliamentary speeches are analyzed and members of parliament (MPs) are the main subjects of study, the full speeches can be defined as the unit of observation. If the same MPs are to be analyzed in newspaper articles, in contrast, it can be helpful to restrict the analysis to the paragraphs mentioning the MPs, leaving paragraphs discussing other topics aside.

Second, a crucial aspect to consider during preprocessing is that most automated text analysis applications are language-specific. If documents in more than one language are to be included in the same analysis, they can either be translated into one language and then analyzed by one single model, or they can be analyzed with separate models. The former has the advantage that one result is estimated that holds for the whole corpus. Some semantic nuances of the texts such as emotions, however, can be lost during the translation. The latter, in contrast, suffers from the problem that the results produced by the different models may not be straightforwardly comparable. Vries, Schoonvelde, and Schumacher (2018) have shown that, for many text-as-data applications, machine translation performs almost as well as expert translation. Therefore, researchers should seriously consider this option when working with multi-lingual corpora. Moreover, the quality of automatic translation is increasing steadily, with services such *DeepL*² being currently the state of the art.

Third, the information in the texts needs to be reduced by removing “stop words” (that is, common words such as “and” or “the”), punctuation, and numbers. The rationale to reduce this information is that not all text elements contain important information for the word distributions used in the estimations. Stemming or lemmatizing is another procedure to reduce complexity. Stemming cuts word endings, while lemmatizing transforms each word into its basic form. For most languages other than English, lemmatization should be preferred since there are many irregular conjugations and declensions. Researchers can also opt to enrich the text data by identifying the part-of-speech (POS) of words (that is, whether a given word is a verb, a noun, etc.), building n-grams (combinations of words), and extracting information on the syntactical dependencies of the words in the texts. Such methods can prove very useful to analyze short texts for which simple word distributions entail not enough variance. Since they add more layers of basic information such as the word order, they can considerably improve the estimations in some scenarios. All these decisions depend on the specific goals of the analysis as well as on the nature of the texts. Punctuation, for example, may only add noise for most estimations, but it has been found to be useful in classification of emotions. Because of this uncertainty, it is generally recommended to consider several alternatives and test empirically whether they improve the estimations.

The preprocessing of the text data has a decisive influence on the results (Denny and Spirling 2017). Therefore, it is recommended to either extensively test the influence of every preprocessing step or, even better, to include the preprocessing parameters in the analysis procedures.

4. METHODS

Text-as-data applications are one of the most dynamic areas of political science methodology. In the last years, contributions from this area have steadily increased. As the range of applications grows, it becomes difficult to keep track of all developments. The subsequent discussion tries to provide a broad overview over the area of text-as-data methods and applications. More precisely, we suggest that the many different applications can be grouped according to three different research goals: extraction of specific information (concept identification), theory-driven allocation (classification), and inductive exploration of the underlying dimensionality (discovery).

4.1 Concept Identification

The goal of concept identification is to find and extract the specific text passage that refers to a concept of interest. Concepts can be highly complex, such as the degree of conflict in a political speech, but also more straightforward, such as the names of governors of US states in press releases. In broad terms, concept identification methods can be separated into dictionary-based approaches and named entity recognition approaches. Applications of dictionaries, also referred to as “ontologies”, “lists”, or “gazetteers” in some literatures, involve matching keywords in the texts of interest. They are sometimes criticized for their simplicity. However, if the operationalization produces a comprehensive set of keywords that can be mapped unambiguously to a concept, such approaches are highly reliable and efficient. A good example are names of politicians or political parties, which are quickly compiled and mostly

refer unequivocally to the actors under concern (see Gilardi and Wüest 2017; Müller 2014; Wüest et al. 2016). Sentiment analysis traditionally was also conducted using dictionaries of word polarities (Young and Soroka 2012). Although these approaches are being increasingly replaced by supervised classifications using machine learning, they keep being actively developed, for instance for applications in multi-lingual settings (Proksch et al. 2019). Technically, the matching of dictionaries can be implemented using regular expressions (regex), which can be implemented using libraries that are part of the base distribution of both R and Python.

A more complex set of methods for concept identification is usually termed Named Entity Recognition (NER). NER approaches are based on machine learning, which means that specific concepts are recognized by a model using linguistic rules and bag-of-words information from the word contexts of these concepts. There are NER tools that are trained on such large corpora such as Wikipedia sites (e.g. the Stanford NER or polyglot library in Python) that they can be applied off the shelf. Hence, no dictionary has to be built when using NER tools, but they are usually only able to detect a restricted set of concepts such as persons, locations, dates, or organizations. However, these are usually the concepts researchers in comparative policy analysis are interested in. The detection of locations, for example, can be used to assign documents to geographical units that are the subject of policy diffusion studies (Ciocan and Wüest 2017; Gilardi, Shipan, and Wüest 2020).

4.2 Classification

Supervised classification tasks can be defined as a separate set of text-as-data methods. Text classification can either be used to assign class memberships, e.g. to which policies the documents of interest can be categorized, or to estimate class shares in documents, e.g. the relative importance of different policy debates in the same documents (Jurafsky and Martin 2009; King, Pan, and Roberts 2013).

The approach is similar for most applications. First, a training set needs to be built, which usually involves the manual coding of a sample of the data to be classified. Increasingly, researchers in political science also use crowd-sourcing to build these training sets (Benoit et al. 2016). Then, a generative model is created and optimized. It uses the hand-coded inputs to calculate the probability that each document belongs to a certain category. Popular algorithms implementing such models are the multinomial naïve Bayes (Conway and White 2012), support vector machines (Meyer 2012), regularized paths for generalized linear models (Friedman, Hastie, and Tibshirani 2010), and maximum entropy (Jurka et al. 2013). Such models can be further optimized using bootstrapped training and cross-validation, evaluating the best trade-off between false positives and false negatives (which is also known as optimization of the receiver operator characteristic, or ROC), and building ensembles, that is, classifiers that include several algorithms or models and perform classification by comparing their predictions.

In many text-as-data applications classifications are necessary first steps in order to compile the corpus of interest, since text data collection may contain a large number of false positives, that is, texts that do not actually belong in the corpus (e.g. Ciocan and Wüest 2017; Gilardi, Shipan, and Wüest 2018; Wüest et al. 2016).

Software tools that allow several kinds of classification tasks are Quanteda, Readme, and RTextTools in R, and especially scikit-learn in Python.

4.3 Discovery

While supervised classification is a deductive exercise in which texts are grouped into categories that are defined theoretically, other approaches are inductive and can be used to discover latent structures in the corpus and situate the texts within this latent structure.

Well-known examples in political science include “Wordfish” (Lowe 2008; Slapin and Proksch 2008), which maps texts onto ideal points on ideological or issue-specific dimensions (Lowe 2013). These methods need very careful text preprocessing, parameter tuning, and testing in order to be reliable.

Another strand of latent variable models are generative mixed-membership models, such as topic models, which can uncover the semantic structure of a corpus (Blei, Ng, and Jordan 2003). Topic models can be a useful tool to identify frames in texts. As DiMaggio, Nag, and Blei (2013: 578, 593) write, “[m]any topics may be viewed as frames ... and employed accordingly ... [T]opic modeling has some decisive advantages for rendering operational the idea of ‘frame’”. In this context, mixed-membership means that these models assume that each document can be assigned to multiple categories, in different proportions. In other words, a given text will not include just one topic, but multiple topics, although different texts will give more or less importance to different topics. A particularly useful variant of topic models is the structural topic model (Roberts et al. 2014; Roberts, Stewart, and Airolidi 2016), which allows the prior distribution of documents and words over topics to be influenced by co-variables. For instance, this allows one to measure how the topics co-vary with time or with other variables of interest. For example, such models can be used to analyze how the topics of newspaper articles vary depending on whether male or female politicians are mentioned (Gilardi, Shipan, and Wüest 2020; Gilardi and Wüest 2017). Furthermore, topic models can also be used to explore corpora in order to uncover novel measures or research questions (see Wüest 2018).

Scaling procedures can be conducted using the packages *austin* or *quanteda* in R, while topic models can be estimated with *gensim* in Python and *stm* in R.

5. NEW DIRECTIONS OF TEXT-AS-DATA APPLICATIONS

We identify three main directions in which text-as-data applications are currently developing. First, causality. Most text-as-data approaches are purely predictive, but social scientists and policy analysis are often interested in establishing causal relationships. Causal inference with text data is not straightforward, but Egami et al. (2018) have put forward a framework to facilitate this task. Egami et al. (2018) suggest estimating the causal effects in sequential experiments. Concretely, the advice is to split the data in a similar way as in the training of supervised classifications. Concretely, one set should be used to optimize the inductive procedures (for example, the number of topics in a topic model), while the other should be used for estimating causal relationships.

Second, another stream of research aims to go beyond the many simplifications conducted when preprocessing the texts, which usually involve discarding a lot of potentially useful information. Computer science research on word embeddings and artificial neural networks, or “deep learning” as it is often referred to, will likely gain prominence in text-as-data applications in political science and policy analysis (LeCun, Bengio, and Hinton 2015; Mikolov et al. 2013).

Third, the “as-data” trend will not stop at text. The next frontier involves images, sounds, and videos (e.g. Dietrich, Enos, and Sen 2019; Joo and Steinert-Threlkeld 2018).

6. CONCLUSION

The goal of this chapter was to offer an overview of text-as-data methods for comparative policy analysis, a field in which these methods have been used less extensively than in political science more generally. These methods have a very high potential to develop new tests of existing theories and to uncover new aspects of policy making that were previously very hard to study. In particular, we identify a number of advantages compared to traditional approaches relying on manual coding.

First, text-as-data approaches are scalable. That is, once the method has been developed or adapted for a specific project, additional material can be analyzed with sharply decreasing marginal costs. This is especially useful for projects that are intended to continue over a long period of time.

Second, text-as-data approaches make it easier to extend projects to new areas. Here, there are still non-trivial start-up costs to adapt existing procedure to new contexts, but the extension can be done much more efficiently than with traditional approaches. Often, new contexts involve new languages. Many text-as-data approaches already work quite well using automatic translation, whose performance is improving quickly.

Third, text-as-data approaches permit retroactive adjustments to data that were previously coded. With manual approaches, this is practically impossible because of the prohibitive costs. Once a large data set has been coded manually, for all intents and purposes it is fixed. But exploring alternative coding strategies can be useful for a number of reasons. First, one could question the original choices on theoretical or substantive grounds. Second, for projects that run over long periods, such as the Comparative Agendas Project, the categories do not necessarily remain constant. New issues emerge, and others lose importance. Adjusting the coding scheme retroactively, to ensure comparability over time, is much easier using automated approaches.

Fourth, automated approaches increase transparency and facilitate replication. Replicating the original coding is prohibitively costly if it was conducted manually. With text-as-data approaches, it is in principle possible to retrace how a data set was constructed, starting with the raw data. Of course, this depends in large part on how carefully researchers document their procedures and make all data available, which might also be a problem for copyright or privacy reasons. However, the bar for replicability is definitely lower when using automated approaches.

Fifth, one might be discouraged by the fact that the methods available today are not quite advanced enough to do something that could be done using manual approaches. However, methods are improving very quickly, and one should focus more on what will be possible in a couple of years than on what can be done right now, keeping in mind the other advantages of automated approaches.

To conclude, we encourage scholars to invest time learning these methods, and the comparative policy analysis community to offer training to make the learning curve less steep. We hope that this chapter is a helpful starting point.

NOTES

1. <https://www.comparativeagendas.net/pages/master-codebook>, accessed March 7, 2018.
2. <https://www.deepl.com/translator>, accessed March 7, 2018.

REFERENCES

- Baumgartner, F. R., Green-Pedersen, C., and Jones, B. D. (2006). Comparative studies of policy agendas. *Journal of European Public Policy* 13 (7): 959–74.
- Baumgartner, F. R. and Jones, B. D. (1993). *Agendas and Instability in American Politics*. Chicago, IL: University of Chicago Press.
- Baumgartner, F. R., Breunig, C., and Grossmann, E. (2019). The Comparative Agendas Project Intellectual Roots and Current Developments. In: Baumgartner, F. R., Breunig, C., and Grossmann, E. (Eds.) *Comparative Policy Agendas: Theory, Tools, Data* Oxford: Oxford University Press.
- Baumgartner, F. R., Jones, B. D., and Wilkerson, J. (2011). Comparative studies of policy dynamics. *Comparative Political Studies* 44 (8): 947–72.
- Benoit, K., Conway, D., Lauderdale, B., Laver, M., and Mikhaylov, S. (2016). Crowd-sourced text analysis: reproducible and agile production of political data. *American Political Science Review* 110 (2): 278–95.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning* 3: 993–1022.
- Boushey, G. (2016). Targeted for diffusion? How the use and acceptance of stereotypes shape the diffusion of criminal justice policy innovations in the American states. *American Political Science Review* 110 (1): 198–214.
- Braun, D. and Gilardi, F. (2006). Taking “Galton’s problem” seriously: towards a theory of policy diffusion. *Journal of Theoretical Politics* 18 (3): 298–322.
- Card, D., Boydston, A. E., Gross, J. H., Resnik, P., and Smith, N. A. (2015). The media frames corpus: annotations of frames across issues. Paper presented at the Annual Meeting of the Association for Computational Linguistics, Beijing, July.
- Card, D., Gross, J., Boydston, A., and Smith, N. A. (2016). Analyzing framing through the casts of characters in the news. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, TX, 1410–20.
- Ciocan, D. and Wüest, B. (2017). How MENA media frame the Arab Spring. Paper presented at the Annual Meeting of the American Political Science Association, San Francisco, CA.
- Collingwood, L. and Wilkerson, J. (2012). Tradeoffs in accuracy and efficiency in supervised learning methods. *Journal of Information Technology & Politics* 9 (3): 298–318.
- Conway, D. and White, J. M. (2012). *Machine Learning for Hackers: Case Studies and Algorithms to Get You Started*. Cambridge, MA: O’Reilly.
- Denny, M. J. and Spirling, A. (2017). Text preprocessing for unsupervised learning: why it matters, when it misleads, and what to do about it. Unpublished manuscript, Pennsylvania State University.
- Dietrich, B. J., Enos, R. D., and Sen, M. (2019). Emotional arousal predicts voting on the US Supreme Court. *Political Analysis* 27 (2): 237–43.
- DiMaggio, P., Nag, M., and Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: application to newspaper coverage of US government arts funding. *Poetics* 41 (6): 570–606.
- Dolowitz, D. P. (2000). Introduction to the special issue on policy transfer. *Governance* 13 (1): 1–4.
- Dolowitz, D. and Marsh, D. (1996). Who learns what from whom: a review of the policy transfer literature. *Political Studies* 44 (2): 343–57.
- Dowding, K., Hindmoor, A., and Martin, A. (2016). The Comparative Policy Agendas Project: theory, measurement and findings. *Journal of Public Policy* 36 (1): 3–25.
- Egami, N., Fong, C. J., Grimmer, J., Roberts, M. E., and Stewart, B. M. (2018). How to make causal inferences using texts. Unpublished manuscript, Princeton University.

- Elder, C. D. and Cobb, R. W. (1984). Agenda-building and the politics of aging. *Policy Studies Journal* 13 (1): 115–29.
- Evans, J. A. and Aceves, P. (2016). Machine translation: mining text for social theory. *Annual Review of Sociology* 42: 21–50.
- Fisher, D. R., Leifeld, P., and Iwaki, Y. (2013). Mapping the ideological networks of American climate politics. *Climatic Change* 116 (3–4): 523–45.
- Fisher, D. R., Waggle, J., and Leifeld, P. (2013). Where does political polarization come from? Locating polarization within the US climate change debate. *American Behavioral Scientist* 57 (1): 70–92.
- Flores, R. D. (2017). Do anti-immigrant laws shape public sentiment? A study of Arizona’s SB 1070 using Twitter data. *American Journal of Sociology* 123 (2): 333–84.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33 (1): 1–22.
- Gerner, D. J., Schrodt, P. A., Francisco, R. A., and Weddle, J. L. (1994). Machine coding of event data using regional and international sources. *International Studies Quarterly* 38 (1): 91–119.
- Gilardi, F., Shipan, C. R., and Wüest, B. (2020). Policy diffusion: the issue-definition stage. University of Zurich and University of Michigan.
- Gilardi, F. and Wüest, B. (2017). Newspaper coverage of female candidates during election campaigns: evidence from a structural topic model. Paper presented at the Annual Meeting of the American Political Science Association, San Francisco, CA.
- Golder, S. A. and Macy, M. W. (2014). Digital footprints: opportunities and challenges for online social research. *Annual Review of Sociology* 40: 129–52.
- González-Bailón, S. (2017). *Decoding the Social World: Data Science and the Unintended Consequences of Communication*. Cambridge, MA: MIT Press.
- Graham, E. R., Shipan, C. R., and Volden, C. (2013). The diffusion of policy diffusion research in political science. *British Journal of Political Science* 43 (3): 673–701.
- Grimmer, J. (2013). Appropriators not position takers: the distorting effects of electoral incentives on Congressional representation. *American Journal of Political Science* 57 (3): 624–42.
- Grimmer, J. and Stewart, B. M. (2013). Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21: 267–97.
- Hopkins, D. and King, G. (2010). A method of automated nonparametric content analysis for social science. *American Journal of Political Science* 54 (1): 229–47.
- Jatana, N., Puri, S., Ahuja, M., Kathuria, I., and Gosain, D. (2012). A survey and comparison of relational and non-relational database. *International Journal of Engineering Research & Technology* 1 (6): 1–5.
- John, P. (2006). The Policy Agendas Project: a review. *Journal of European Public Policy* 13 (7): 975–86.
- Jones, B. D. and Baumgartner, F. R. (2005). *The Politics of Attention: How Government Prioritizes Problems*. Chicago, IL: University of Chicago Press.
- Joo, J. and Steinert-Threlkeld, Z. C. (2018). Image as data: automated visual content analysis for political science. Preprint arXiv:1810.01544.
- Jungherr, A. and Theocharis, Y. (2017). The empiricist’s challenge: asking meaningful questions in political science in the age of big data. *Journal of Information Technology & Politics* 14: 97–109.
- Jurafsky, D. and Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ: Prentice Hall.
- Jurka, T. P., Collingwood, L., Boydston, A. E., Grossman, E., and van Atteveldt, W. (2013). RTextTools: a supervised learning package for text classification. *The R Journal* 5 (1): 6–12.
- King, G., Pan, J., and Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review* 107 (2): 326–43.
- Klüver, H. (2013). Lobbying as a collective enterprise: winners and losers of policy formulation in the European Union. *Journal of European Public Policy* 20 (1): 59–76.
- Klüver, H. and Mahoney, C. (2015). Measuring interest group framing strategies in public policy debates. *Journal of Public Policy* 35 (2): 223–44.
- Klüver, H., Mahoney, C., and Opper, M. (2015). Framing in context: how interest groups employ framing to lobby the European Commission. *Journal of European Public Policy* 22 (4): 481–98.
- Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage.

- Laver, M., Benoit, K., and Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review* 97 (2): 311–31.
- Lazer, D. and Radford, J. (2017). Data ex machina: introduction to big data. *Annual Review of Sociology* 43: 19–39.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521: 436–44.
- Leifeld, P. (2013). Reconceptualizing major policy change in the Advocacy Coalition Framework: a discourse network analysis of German pension politics. *Policy Studies Journal* 41 (1): 169–98.
- Leifeld, P. and Haunss, S. (2012). Political discourse networks and the conflict over software patents in Europe. *European Journal of Political Research* 51 (3): 382–409.
- Linder, F., Desmarais, B. A., Burgess, M., and Giraudy, E. (2018). Text as policy: measuring policy similarity through bill text reuse. *Policy Studies Journal*, forthcoming.
- Loftis, M. W. and Mortensen, P. B. (2018). Collaborating with the machines: a hybrid method for classifying policy documents. *Policy Studies Journal*, forthcoming.
- Lowe, W. (2008). Understanding wordscores. *Political Analysis* 16 (4): 356–71.
- Lowe, W. (2013). Putting it all on the line: some unified theory for text scaling. Paper prepared for the American Political Science Association meeting, Chicago, IL, September.
- Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., and Tingley, D. (2015). Computer-assisted text analysis for comparative politics. *Political Analysis* 23 (2): 254–77.
- Meyer, D. (2012). Support vector machines. Technische Universität Wien, Austria.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. CoRR, abs/1301.3781.
- Monroe, B. L. and Schrod, P. A. (2008). Introduction to the special issue: the statistical analysis of political text. *Political Analysis* 16 (4): 351–5.
- Müller, L. (2014). *Comparing Mass Media in Established Democracies: Patterns of Media Performance*. Basingstoke: Palgrave Macmillan.
- Nowlin, M. C. (2016). Modeling issue definitions using quantitative text analysis. *Policy Studies Journal* 44 (3): 309–31.
- Pierson, P. (1993). When effect becomes cause: policy feedback and political change. *World Politics* 45: 595–628.
- Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review* 94 (2): 251–67.
- Proksch, S.-O., Lowe, W., Wäckerle, J., and Soroka, S. (2019). Multilingual sentiment analysis: a new approach to measuring conflict in legislative speeches. *Legislative Studies Quarterly* 44 (1): 97–131.
- Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., and Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Political Science Review* 54 (1): 209–28.
- Roberts, M. E., Stewart, B. M., and Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association* 111 (515): 988–1003.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S., Albertson, B., and Rand, D. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science* 58: 1064–82.
- Sebók, M. and Berki, T. (2017). Incrementalism and punctuated equilibrium in Hungarian budgeting (1991–2013). *Journal of Public Budgeting, Accounting & Financial Management* 29 (2): 151–80.
- Simmons, B., Dobbin, F., and Garrett, G. (2006). Introduction: the international diffusion of liberalism. *International Organization* 60 (4): 781–810.
- Slapin, J. B., and Proksch, S.-O. (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science* 52 (3): 705–22.
- Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., and Hartgerink, C. H. J. (2016). The academic, economic and societal impacts of open access: an evidence-based review. *F1000Research* 5 (632).
- Vries, E. de, Schoonvelde, M., and Schumacher, G. (2018). No longer lost in translation: evidence that Google translate works for comparative bag-of-words text applications. *Political Analysis* 26 (4): 417–30.
- Walker, J. L. (1969). The diffusion of innovations among the American states. *American Political Science Review* 63 (3): 880–899.

- Wilkerson, J. and Casas, A. (2017). Large-scale computerized text analysis in political science: opportunities and challenges. *Annual Review of Political Science* 20: 529–44.
- Wilkerson, J., Smith, D., and Stramp, N. (2015). Tracing the flow of policy ideas in legislatures: a text reuse approach. *American Journal of Political Science* 59 (4): 943–56.
- Wüest, B. (2018). Selective attention and the information environment: citizens' perceptions of political problems in the Swiss federal election campaign 2015. Paper proposal for the Special Issue of the *Swiss Political Science Review* on the 2015 Swiss national election.
- Wüest, B., Bütikofer, S., van der Lek, A., and Gantenbein, F. (2016). *Selects Media Analyses 2015: Election Campaign in Swiss National Media. Codebook & Technical Report*. Zurich: University of Zurich – Selects – FORS.
- Young, L. and Soroka, S. (2012). Affective news: the automated coding of sentiment in political texts. *Political Communication* 29 (2): 205–31.

PART IV

MIXED METHODS AND MULTI-METHODS

13. Critical multiplism for comparative policy analysis

William N. Dunn and B. Guy Peters

1. INTRODUCTION

Comparative policy analysis, and indeed any systematic process of comparative inquiry, should attempt to balance and deliberately relate two forms of inquiry. First, we should employ the comparative method, which implies cross-case analysis. As Arend Lijphart pointed out in his seminal article (1971; see Peters, Chapter 2 this volume), the comparative method depends heavily upon the reasoned selection of cases and the comparison of findings across those cases. The process of comparison generally involves comparing relations across cases by identifying and analyzing a dependent variable and one or more independent variables. In this context, large-N studies using statistical methods typically investigate the relations of one or more Xs with a Y across a set of available cases, cases whose internal characteristics are partially or wholly unknown. For example, we may regress values of the Gini coefficient on per capita GDP for a sample of countries in the absence of within-case knowledge about differences in taxation policies that are likely to affect the distribution of income.

Hence, in addition to independent variables within a regression equation, it is important to acquire an understanding of properties of processes and structures within individual cases. This can be done by identifying intervening variables in a data set or by employing process-tracing or other forms of case study design and analysis (see Fontaine, Chapter 16 this volume; and Yin 2017). While cross-case analysis may estimate one or more presumed causal relations among units included in the regression equation, within-case analysis may identify causal patterns by examining decision processes and outcomes (but see Beck 2010). Without within-case analysis the researcher often presumes patterns of causation among units without exploring units from within (see Gordon and Smith 2004).

Overall, however, large-N comparative political research has its limitations. Although different forms of large-N regression analysis are readily available in econometrics texts and handbooks (e.g. Kennedy 2003), quantification does not easily confer benefits to mixed-methods researchers who recognize the importance of understanding the contexts of countries included in quantitative comparative analysis (see Pollitt 2013). Kasza (2005: 425), for example, argues that “the more cases scholars include in a study, the less likely they are to grasp the full range of information pertinent to understanding those cases”. In this context, it has even appeared to some (Ahmed and Sil 2012) that that the task of combining quantitative and qualitative methods is so daunting that we would be wise to specialize in one method or the other if young scholars wish to achieve the methodological mastery required to gain professional respect and rewards including tenure and promotion.

That said, quantitative analysis is a powerful means for investigating causality, even in “observational” studies that depart from canons of experimental design (see Angrist and Pischke 2008, 2014). Nevertheless, unknown properties of cases in a time-series or

cross-sectional analysis are often regarded as “uncountable” aspects of politics and policy. Large-N studies are those most prone to exclude within-case properties which, because they are presumed to be uncountable, cannot be related to quantitative studies that presume “countability”. Indeed, the large number of large-N quantitative studies appears to be inversely related to within-case qualitative studies, such that the validity-diminishing effects of unknown or omitted variables within cases is a source of threats to validity, one which at best may be revealed only by creating normal probability plots, measuring heteroscedasticity, or estimating standard errors. In effect, there is virtually no way to open up these “within-case” black boxes unless conventional large-N cross-case analysis – which excels at supplying measures of average effects of one or more independent variables but ignores context – is supplemented by within-case analyses that permit the investigation of what goes on within the black boxes. Until relatively recently, many comparativists have been forced to abandon mixed methods that deliberately combine cross-case and within-case analysis.

2. CRITICAL MULTIPLISM AND MIXED METHODS

A promising approach to mixed-methods research is what Cook (1985), Houts, Cook, and Shadish (1986), Shadish, Cook, and Campbell (2002: 460), and Shadish, Cook, and Houts (1986) call critical multiplism. Critical multiplism, which is based on Roy Bhaskar’s *critical realism* (Bhaskar 1975, 1979), works from a social ontology that presumes at a philosophical level that there is an external social reality, while acknowledging that social science research methods can at best approximate that presumed social reality. The critical realism that justifies critical multiplism, in contrast to logical positivism, avoids what Bhaskar has pointedly characterized as the “epistemic fallacy”. The epistemic fallacy conflates questions about the existence of the world with questions of whether we can prove that the world exists. The social world is presumed to exist, but we cannot prove it, let alone obtain absolute knowledge of its properties, structures, and processes.

Critical multiplism contends that all methods are fallible; none can be regarded as unequivocally superior to others. Singly or together, the quantitative and qualitative methods that figure prominently in “mixed-methods” research are imperfect and fallible, but corrigible. The basic proposition of critical multiplism is that a choice of a single method, qualitative or quantitative, cannot be justified. Hence, whenever practicable two or more different methods should be selected in order to triangulate on a common object, with the aim of producing an optimally valid inference.¹ Cook (1985), Shadish and Cook (1999), and Flick (2007) emphasize the imperfect nature of all methods for probing what is known and what is worth knowing.

It is now not easy to assume that one is describing a social world that is lawfully fixed, deterministically ordered, and can be perfectly described with elegant and simple functional relationships ... to assume that everything of importance can be measured ... that value free measurement is possible ... that our theories are perfectly specified. (Cook 1985: 29)

In essence, critical multiplism is a form of multiple triangulation grounded in critical realism and based on mixed methods. The term “mixed” usually means that two or more methods must be a heterogeneous combination of quantitative and qualitative methods (Figure 13.1). The process of triangulation, even with homogeneous combinations of methods, however, may

help create optimally plausible but fallible causal claims about the single and joint effects on a focal policy outcome of homogeneous sets of quantitative or qualitative variables.

The ultimate aim of critical multiplism is producing optimally plausible causal inferences, of which there are several types (see Collingwood 1940: 285–300; Cook and Campbell 1979: 25–7; von Wright 1971). The most important for mixed-methods research are: (1) cause as intention, purpose, or reason, for example, human dignity as a cause of drafting legislation on human rights. Although some contend that qualitative methods cannot be causal, the bulk of scholarly opinion concludes that qualitative data are causally relevant and “real” (Maxwell 1992; Miles and Huberman 1994); (2) cause as a practical action or manipulation of human and material resources to achieve a valued outcome, as in an experiment in the most general sense promoted by John Dewey’s experimentalism (Woodward 2003; also see Dunn 2019); and (3) cause as part (along with effect) in a causal mechanism, for example, a light switch. Other less important definitions of cause are (4) cause as a set of initial and ensuing conditions explained by a law, as in a deductive-nomological (D-N) explanation such as Marshall’s law of diminishing utility of money; and (5) cause as a sequence of measured operations learned by rote memory, for example, a recipe the effect of which is a meal (Gasking 1955). Recipes require no scientific explanation, unless one consults a physics cookbook that explains the interaction of chemical constituents of a cake. Finally, (6) a deflationary cause is a “cause” reduced to linguistic expressions and manners of speaking that express nothing of causal relevance (Horwich 1990).

Cause has also been defined in terms of necessary and sufficient conditions, but this simple notion of cause has been discarded by scholars including Mackie (1984: 62; see also Shadish, Cook, and Campbell 2002: 4–5), who acknowledges the complexity of causation by defining a cause as “an *insufficient* but *nonredundant* part of an *unnecessary* but *sufficient* condition” (italics original).

Pearl and Mackenzie (2018) provide a comprehensive overview of different meanings of causation (see also Glymour 2001; Pearl, Glymour, and Jewell 2016), but only after reaching the compelling conclusion that classic statistics, from simple measures of association and correlation to regression analysis and econometrics, provides no means of identifying, depicting, or communicating causality. Until relatively recently, statistics has been a causation-free language and discipline epitomized by the oft-heard mantra: “Correlation is not causation”.

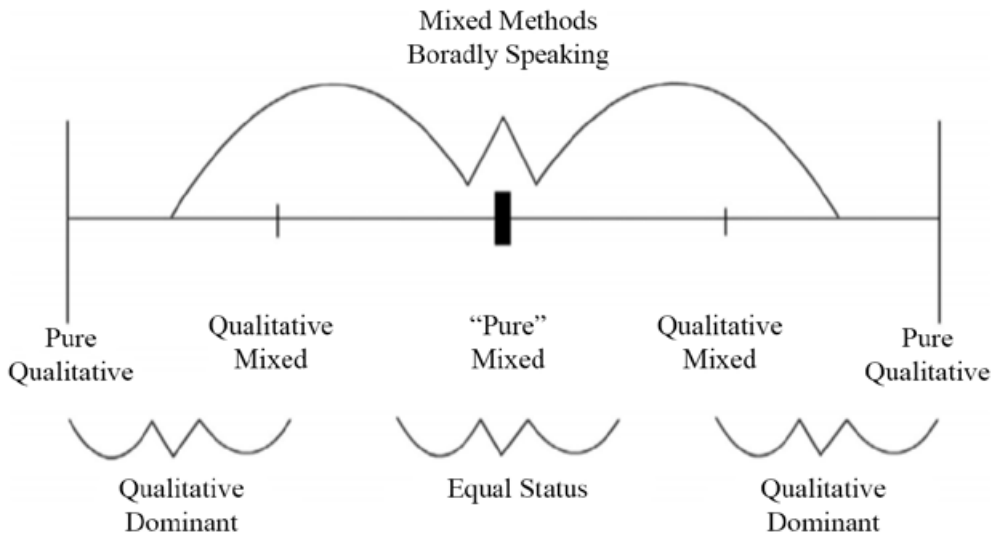
Mixed methods are “mixed” because they employ qualitative as well as quantitative data to describe or explain a focal policy outcome. By contrast, the use of multiple methods does not carry the requirement that qualitative and quantitative methods be used at the same time. Multiple methods only require that two or more quantitative (or qualitative) methods be used, for example, when standardized achievement tests provide the data for regressing STEM exam scores on per capita educational expenditures in a large-N cross-national study. Similarly, qualitative multiple methods of participant observation and process-tracing may be used to trace the inadequate implementation of a policy authorizing the expenditure of funds by a Ministry of Education earmarked for secondary and elementary schools in rural areas, along with the reasons offered by policy makers for the shortfall. Appropriate examples of qualitative multiple methods may be found in volumes such as Miles and Huberman (1994). Similarly, but in the quantitative-by-qualitative mixed-methods domain, regression analysis may be employed in a large-N study to estimate the effects of rapid economic growth on civil unrest, followed by a process-tracing study of civic disorder in one or several of the modeled countries.²

2.1 Multiple Operationism

An exhaustive range of mixed- and multiple methods is displayed in Figure 13.1. One of these is multiple operationism, which involves the use of two or more methods to operationalize the definition of a construct or variable, that is, specify two or more operations for observing and measuring the same thing – for example, a standardized scale of political efficacy and qualitative interviews to assess whether subjects meant the same thing when they gave a common response such as “strongly agree”. Multiple operationism stands in sharp contrast to the use of a single operational definition that implies that the operational definition, despite its unknown biases, is a direct measure of some ontological property. Here, the error is the “epistemic fallacy”, which presumes that through our single operational definition we are measuring ontological properties themselves, thus violating a central tenet of critical realism: “Intelligence is what an intelligence test measures”.

Using ethnographic interviews and participant observation qualify as multiple operationism, as do multiple measures of standardized scales of political efficacy and alienation and their use in linear and nonlinear regression equations. The use of ethnographic interviews with regression analysis also qualifies. Nested analysis (Lieberman 2005: 436), while more general in scope, assumes an interest both in exploring general relationships and explaining individual cases. Nested analysis might include any combination of methods displayed in Figure 13.1, all or most of which might serve as a check on a range of reactive measures and biases originating in data archives, researchers, respondents, and research instruments (Webb et al. 1999).

Lieberman (2005) provides opportunities for creatively employing mixed methods in many research settings. For example, “a nested research design implies that scholars will pose questions in forms such as ‘What causes social revolutions?’ while simultaneously asking



Source: Johnson, Onwuegbuzie, and Turner (2007: 124).

Figure 13.1 *Range of types of mixed methods*

questions such as ‘What was the cause of social revolution in France? Nested analysis helps scholars to ask good questions when analyzing their data and to be resourceful in finding answers’ (Lieberman 2005: 436).

Nested analysis thus conforms “broadly speaking” (see the top of Figure 13.1), to mixed methods. Measures used in regression equations may be selected for convenience from readily available data sets, or because cases suit a researcher’s interests. Convenience may be acceptable up to a point, but by having more detailed understanding of the causal processes coming from within-case analysis it may be possible to refine the process of measurement and improve the overall scientific quality of the research (Adcock and Collier 2001).

2.2 Multiple Observers

The perspectives of two or more observers help generate multiple plausible rival hypotheses based on convergent and divergent perspectives of a focal context. Triangulation occurs through observers whose biased perspectives have been conditioned by different norms, values, and taken-for-granted practices of their own general and academic cultures. However, in the usual one-observer, one-country case study it is unclear whether observations are based on the constructs of the observer or of the object observed. This problem, which Theodore Abel (1948; but see Abel 1975) long ago characterized as the conflict between the “imaginative reconstruction” of cases and the operation of *Verstehen* (understanding), is alleviated today by ethnographic and related methods including Q-methodology (Brown 1993) and the semantic differential (Osgood, Suci, and Tannenbaum 1957). These techniques, like all others, are subject to error and biases, the most important of which is likely the social desirability bias that accompanies the responses from subjects who wish to please interviewers. Procedures for identifying and mitigating the effects of this and other biases are available and well known (Webb et al. 1999).

2.3 Multiple Theories and Construct Validation

One of the advantages of triangulating with two or more theories is to establish the validity of theoretical constructs such as social capital, democracy, and governance. Among the quantitative measures of the construct of governance are the Worldwide Governance Indicators, which rank countries on six aspects of “good governance”. Thomas (2010) cautions that problems of bias and lack of comparability raise questions about the utility of these indicators and the extent to which they meet standards of construct validity, that is, the composite measure of “good governance” should converge, or correlate, with free elections, while it should discriminate, or correlate negatively, with an index of press freedom. In one of the first construct validation studies of its kind, Thomas (2010) finds that the construct of governance, as measured by the World Governance Indicators, has questionable construct validity.

2.4 Meta-Analysis and Meta-Regression Analysis

Methods for conducting systematic reviews, meta-analyses, and meta-regression analyses are becoming essential tools for the social sciences and comparative policy analysis. Although meta-analyses in the area of comparative politics are thus far relatively rare, one of the authors used the Advanced Search Utility of Google Scholar to identify some 1,371 country and

comparative evaluations of civic education, social engagement, and political participation published in the period 2010–2015.³ A cross-case analysis of 40 of these studies showed that all used some form of experimental design (randomized experiment, quasi-experiment, natural experiment), with 10 showing practically significant effect sizes ($d = 0.20$ or larger). The remaining 30 showed no practically significant effect ($d < 0.20$) or insufficient information to reach a conclusion.

At the same time, mixed-method statistical and ethnographic analyses of several cases were conducted. Many of the cross-case findings suggested that civic education programs that focus on democratic values and institutions were effective, but within-case ethnographic analyses suggest that young people possessed a “sense of civic efficacy [that] interacts with their interpretations of historical injustice and the civic messages mediated by teachers, families, peers, and communities” and that “Young people struggle to define and enact appropriate civic action, at times working outside unjust systems as a means of fostering change” (Bellino 2015: 118). The acquisition of knowledge about democracy appears to induce many to recognize the injustice of current politics and, consequently, see *non-voting* as a civic duty.

Li, Owen, and Mitchell (2018) examine a large body of research on the relation between democracy and foreign direct investment (FDI). The authors perform a meta-regression analysis to test whether research on this relation is consistent with observable evidence from previous studies. After using meta-regression analysis to examine 229 model estimates from 40 prior meta-analyses,⁴ they find publication bias (i.e. potentially biased conclusions) suggesting that published research in scientific journals, as distinguished from “grey literature” in unpublished papers and research reports of think tanks and NGOs, present causal models with different dependent variables. Coefficients and standard errors understate the strength of the relation between democracy and FDI. Their analysis also shows that research in this area presents contradictory findings.

The preceding discussion suggests, among other things, a number of common but faulty beliefs among mixed-methods researchers. First, there is a conflation of sample size, on one hand, and the nature of the methodologies employed to make observations. First, small-N research, because it typically uses some sort of case study methodology, is not necessarily “qualitative” at all, except in the somewhat vulgar statistical sense of discrete or nominal measurement – indeed, most case studies might better be termed quasi-qualitative, as they are not concerned with the interpretive understanding of stakeholders. Second, in the tradition of Vico, Heidegger, Weber, and the “cultural sciences”, the term *qualitative* refers to research that seeks to elicit the subjective understanding of act meanings of those we wish to understand – irrespective of sample size. It is instructive to consider that a large number of cases in the Human Relations Area File (HRAF) were coded and analyzed by Stanley Udy (1959), who used qualitative analysis to test the theory that Weberian bureaucracies are more “pervaded” by their external environments than “pervasive”.

Second, there is not necessarily a conflict between quantitative measurement and qualitative observations, in the tradition of qualitative that refers to act meanings. The title and contents of the book by Osgood, Suci, and Tannenbaum, *The Measurement of Meaning* (1957), provide explicit statements of the complementarity of quantitative and qualitative methods. Once meanings are elicited, they may be placed within matrices, represented as frequency distributions, and analyzed with various measures of Euclidean distance, correlation, and association (see Brown 1993). Some early large-N qualitative studies (Dunn and Swierczek 1977; Yin and

Heald 1975), which use cross-case and within-case analysis, have been followed by numerous case surveys that are based on large-N samples.

Finally, once we give up the questionable notion that external validity requires p-values (see Campbell 1975; Ziliak and McCloskey 2008), case studies must meet equally or more demanding *non*-quantitative standards of “proximal similarity” in order to achieve external validity or generalizability (Campbell 1986; Shadish, Cook, and Campbell 2002: 359–61). Small-N studies thus may have external validity, a term that Campbell recommends we replace with the *principle of proximal similarity*.⁵

3. SOME MIXED-METHODS DESIGNS

The basic principle of mixed-methods designs is the concurrent analysis of two or more cross-case and within-case properties of objects that are quantitative, qualitative, or both. These concurrent analyses are presented below in several mixed-methods designs.

3.1 Nested and Crossed Quantitative Designs

One of the origins of within-case and cross-case analysis is the work of Yin (2017), an experimental psychologist who in the 1970s employed an analysis of variance (ANOVA) paradigm for the analysis of laboratory experiments at MIT. However, Yin soon turned to qualitative analysis to deal with the complexity of urban policy settings. One consequence was the development of cross-case and within-case analysis to analyze large sets of urban policies in complex urban settings (see Yin and Heald 1975).

One approach to nested analysis is an extension of factorial analysis of variance methods pioneered by Sir Ronald Fisher (Fisher 1949) and adapted in “nested” analysis of variance methods used in the design of experiments (Kirk 1982: 455–6). Figure 13.2 displays a nested statistical analysis that involves the partitioning of one level of the independent or policy variable in a nested factorial analysis of variance.

If the dependent variable were total energy consumption in BTUs (British thermal units) and the independent variable were a new energy conservation policy in several countries and its absence in several others, the partition of a “nest” of non-adjacent countries would help isolate the effects of policy diffusion, also known as Galton’s problem. Nesting might also be accomplished by partitioning a “nest” of warm Mediterranean climates and excluding colder Continental climates in which energy consumption may be expected to dampen the effects of energy conservation policy. The impact of the policy on energy consumption would be expected to be larger in the nested case than in the case of a full factorial design.⁶

Linking within-case and cross-case analysis can also be done in other ways. One is by using “nested analysis” (Figure 13.2), an approach advocated by Lieberman (2005), in a general and metaphorical way. In fact, the nesting metaphor is just that, a way to talk about contextual analysis by likening it to a nest within a larger entity such as a tree. However, the value of the metaphor is to communicate the idea that quantitative analysis is not sufficient unless the quantitative variables(s) have substantive meaning. We need to know the meaning or context of an entity before we understand it. For example, unless one understands the nest, the equation of which the nest is part is at best incomplete and at worst meaningless or misleading. This compositional causal strategy, proceeding from within the case to the larger case itself, is the

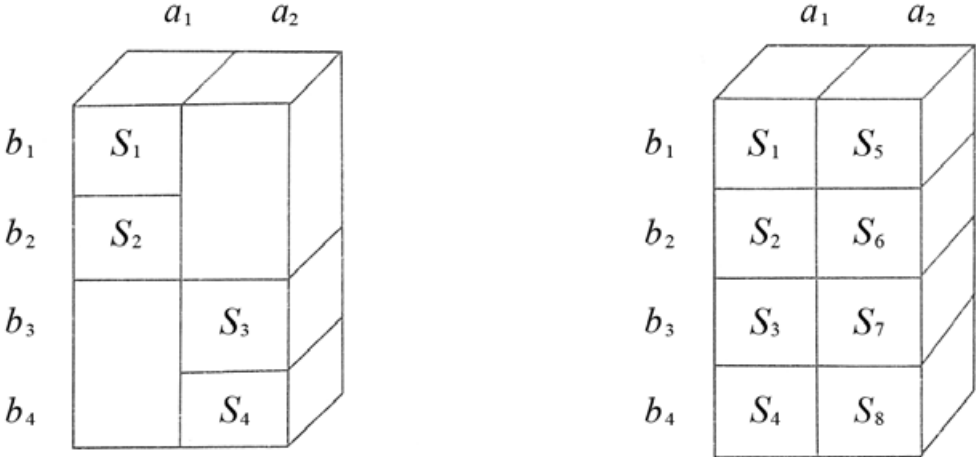
reverse of the decompositional causal strategy, which proceeds from the larger case or cases to the components within the case. Both strategies were originally suggested by Max Weber (e.g. Weber 1946: 204–14; see also Wallace 1971: 102–4).

3.2 Nested Qualitative Designs

The conventional logic of the social sciences, to the extent that multi-method research is being done, is a focus on hypothesis generation through qualitative methods and hypothesis testing through large-N quantitative methods – a resurrection of the logical distinction between the context of discovery and the logic of confirmation (for an example, see King, Keohane, and Verba 1994). Some scholars, however, have argued for the opposite approach, that is, focus on hypothesis generation through quantitative methods, and testing by means of qualitative methods (see Mahoney 2007). The main argument against both sets of scholars is that, while it may be possible to discern patterns of association using large-N comparative research, that does not address causation within the cases. In that style of research there is little or no opportunity to understand the internal causal dynamics of the cases.

Lieberman (2005) has argued that one way to integrate large-N and small-N research, and also quantitative and qualitative analysis, is through his general vision of nested analysis. This form of analysis involves an iterative process of going between large-N analysis – especially regression analysis – and small-N analysis. In the model presented by Lieberman, the process begins with a large-N analysis to identify causally relevant variables and their role in the hypothesized relationship between variables. That cross-case analysis also is used to identify an appropriate set of cases for small-N research by means of within-case analysis.

In contrast to the statistical partitioning of “nests”, Lieberman employs quantitative methods to examine relationships among variables in large-N analysis, followed by the identification through quantitative analysis of typical or atypical cases, for example, those lying at points along a regression line or outliers positioned off the line (see Figure 13.2).



Source: Kirk (1982: 457).

Figure 13.2 *Nested and crossed quantitative designs*

Table 13.1 Experimental binary effect size display

Climate	Energy conservation	Energy conservation
	Program successful	Program unsuccessful
Mediterranean	12 (0.71)	10 (0.40)
Continental	5 (0.29)	15 (0.60)
	17 (100.0)	25 (100.0)

N = 42
 BESD = 0.71 – 0.40 = 0.31

Brunner (1986), after showing how regression analysis produces errors of interpretation when used to do between-case analyses of poverty, recommends a case-wise strategy to poverty. Regression analysis might be used to do a between-case analysis of the responsiveness of poverty levels to independent variables (education, employment) and then write narrative descriptions of cases of poverty located along the regression line. This communicates in the ordinary language of averages and slopes to policy makers with little or no knowledge of statistics. The same procedure may be used to describe cases that fall far off the regression line, for example, outlying high-income households with low levels of education. This procedure has the advantage of potentially redefining poverty and how to measure it, as well as communicating the results of between-case analysis and within-case analysis to policy makers and other stakeholders.

Other scholars have argued for a similar relation between case methods and quantitative methods, albeit beginning within the cases to identify potential causal relationships that can be tested with large-N analysis. This is a variation of the Weberian compositional strategy discussed above. One question which arises in this form of nested analysis is which cases to select for more detailed, within-case analysis. A common recommendation is to select cases that are outliers from the average effect summarized by the regression line. The assumption is that these are the most interesting cases and may reveal something important about the relationships among variables. At the same time, however, the outliers may be just that and say very little about the average effect among variables included in the analysis. Conversely focusing on “typical cases” that sit on the regression line may provide only a limited understanding of causal dynamics. In terms of examining comparative public policy these alternative choices of cases may be especially significant in shaping research findings. Although context matters in many social science studies (Pollitt 2013), it may be especially important for policy studies. The outcomes or effects of policies may well depend on a wide range of political, social and economic factors in an envioning policy process, so that identifying the contextual factors may be crucial for understanding presumed causal relationships. Thus, an outlier from the regression line may be strongly influenced by contextual factors and therefore offer little knowledge about the causal process that is central.

3.3 Experimental Binary Effect Size Display

The binary effect size display is appropriate for cases where we wish to assess the effects on a binary attribute dependent variable of another binary variable that is experimental and presumably causal. For example, cases of 17 countries in which an experimental energy conservation program was implemented successfully and 25 countries where there was no success can be cross-partitioned with countries that have Mediterranean and Continental climates. In this

hypothetical case (Table 13.1) the binary effect size is 0.31, assuming that 42 cases are distributed as in Table 13.1. The BESD (Binary Effect Size Display), which is an appropriate effect size for the analysis of a small set of cases, is equivalent to a (Pearson) correlation coefficient for 2×2 tables. As expected, the conservation program was more successful in Mediterranean climates with a small range of annual temperatures than in Continental climates.

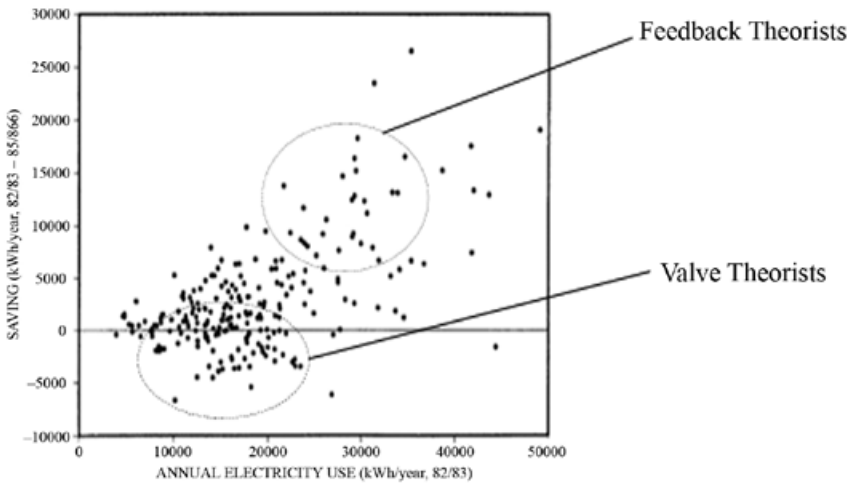
3.4 Qualitative Regression Design

The Hood River Energy Experiment in Oregon employed multistage random sampling to identify a pool of homes, which were randomly assigned to experimental and control groups (see Hirst, Goeltz, and Trumble 1987; Peach et al. 1984). Although there were statistically and practically significant differences between the two groups, there were also outliers from the regression line which departed from requirements of homoscedasticity and linearity, thus deflating measures of correlation and goodness-of-fit. Outlying homes and those along the regression line were studied by energy anthropologists who discovered two kinds of folk theories among persons responsible for changing thermostats within homes (see Kempton 1986).

The first folk theory was that of “valve theorists”, who turned thermostats up or down when homes were cold (or hot), while the second folk theory was that of “feedback theorists”, who set thermostats at a specific level and made small adjustments of one or two degrees from time to time. In the words of one of the energy anthropologists, there are “two theories people use to understand and adjust their thermostats ... the feedback theory and the valve theory. The valve theory is inconsistent with engineering knowledge about the optimal use of energy, but is estimated to be held by 25% to 50% of Americans. Predictions of each of the theories were compared with the operations normally performed in home heat control” (Kempton 1986: 75). The approximate prediction was that better thermostat management may result in savings of billions of dollars annually. A categorical (dummy) variable to capture results of ethnographic interviews and process-tracing can be added to a regression equation, resulting in a substantial improvement in explained variance, lower standard errors, and more accurate predictions of energy usage. Figure 13.3 shows average energy usage (the horizontal line) and energy savings in kilowatt hours above and below the average. If valve theorists below the line were to become feedback theorists, substantial savings would occur. A closer examination of outliers among feedback theorists above the line might uncover within-case factors (e.g. damaged insulation) that would improve savings. Because the basis for the categorical variable in the regression equation is within-case qualitative analysis, we might call this mixed-method qualitative regression analysis (see Figure 13.3). Qualitative regression analysis is a broadly generalizable mixed-method that is applicable in almost any situation where understanding the meanings of quantitative variables affects the validity of claims about outcomes. Note that the qualitative meaning of the categorical (dummy) variable is understood *before* its effects are measured.

3.5 Multiple Observer Comparative and Reflexive Case Study Design

The comparative and reflexive case study design (Campbell 1975) draws on the multiple theoretical and methodological perspectives of two or more researchers who conduct cross-case and within-case analyses of policy processes, structures, and outcomes in other countries and settings (comparative case studies 1–6) as well as their own (reflexive case studies 1 and



Sources: Hirst, Goeltz, and Trumble (1987: 31); Kempton (1986).

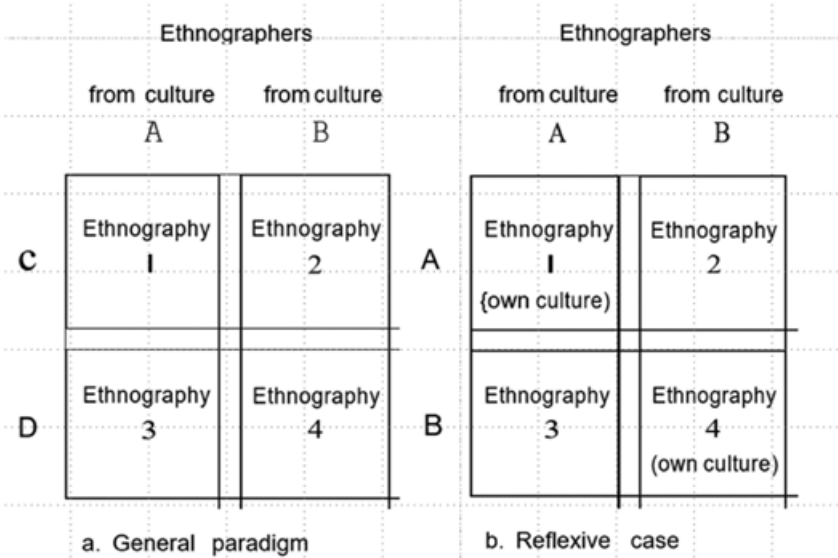
Figure 13.3 *Qualitative regression design: annual electricity use by hypothetical “valve” and “feedback” theorists*

2) (Figure 13.4). The six comparative case studies permit cross-case comparisons by A and B, while the two reflexive case studies by A and B of their own countries or settings enable within-case analyses of the countries in which A and B have lived, worked, studied, and so forth. Both the researchers, A and B, and the objects of investigation – countries or settings A, B, C, and D – presumably contribute jointly to eventual causal claims. With only one researcher-observer, say A, it is impossible to separate subjective judgments and the “facts” of the situation in cases C and D; by adding researcher B, a better assessment of error and bias may be made. This requires both multiple observers (methods) and multiple objects of study.

In the general mixed-methods design (Figure 13.4(a)), two researchers from different cultures study cultures which are not their own. Of the four resultant case studies, the common properties across cases 1 and 3 not shared with 2 and 4 might be attributed to observer A, while the common properties of 2 and 4 not elsewhere present might be attributed to observer B. Looking at row commonalities, the common properties in cases 1 and 2 not present in 3 and 4 could be attributed to culture C, if the analyses of A and B converge (of course they may diverge in some respects and converge in others). Campbell (1975: 191–2) cautions that:⁷

Attributes common to all four [cases] ... are inherently ambiguous, interpretable as either shared biases on the part of the [researchers] ... or shared culture on the part of the societies studied. Note the desirability in this regard of comparing [researchers] ... with as widely differing cultural backgrounds as possible. Insofar as the [researchers] ... come from the same culture, the replication of results becomes more a matter of reliability than validity, as these terms are commonly understood.

The multiple triangulation case study design becomes even more salient when we consider the common practice of using available quantitative data for large-N cross-case analysis. For example, even if researchers A and B studied cases C and D as elements of a larger quantitative



Source: Campbell (1975: 189).

Figure 13.4 Multiple triangulation case study design: comparative and reflexive case studies

data set, it is only through the use of multiple observers that biased interpretations become transparent, particularly those that originate in qualitative understandings (and misunderstandings) of the data that originate in within-case, reflexive case studies. In complex studies involving the cross-case and within-case analysis of revolutions, for example, Skocpol’s *States and Social Revolutions: A Comparative Analysis of France, Russia and China* (1979), it is likely that the multiple triangulation case study design may have produced different causal claims.

3.6 Visual Mixed-Method Designs to Display Multiple Populations, Times, and Settings

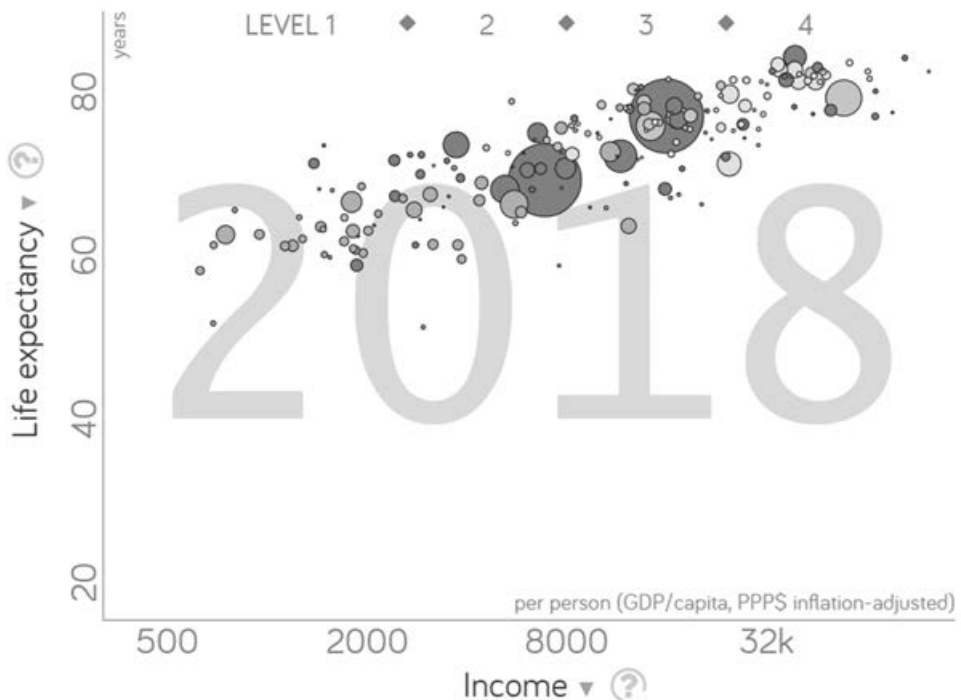
The use of analytic and visual software may be used to examine multiple populations, times, and settings, combining cross-case and within-case quantitative and qualitative analysis. Figure 13.5 shows the relation between life expectancy and per capita GDP for a sample of countries in 2018. The Gapminder software can be extended to include multiple points of time in an extended time-series. The visual display of cross-case and within-case analysis is frequently a better medium of communication than statistical representations with equations expressed in specialized statistical notation (Tuft 2001).

3.7 Bayesian Designs

The final approach to bringing together findings from within-case research in order to make cross-case comparisons relies on the logic of Bayesian statistics (see Gelman et al. 2013;

Humphreys and Jacobs 2015). In particular, this approach examines the extent to which adding information from one strand of research will improve capacities to predict. The standard Bayesian question is how likely we would be to observe a particular set of findings if our hypothesis were true, compared to how likely these findings would be were the hypothesis false.

As a means of combining information gained from qualitative, within-case analysis with information from cross-case analysis, the Bayesian approach asks to what extent our confidence in the findings from a quantitative analysis can be strengthened by the evidence gained from qualitative methods. The example used by Humphreys and Jacobs (2015; see also Bennett 2008) is the use of process-tracing techniques as the qualitative method. This is perhaps the strongest of the possible qualitative methods for these purposes, but in principle other qualitative methods could also be used as a means of strengthening confidence. The procedure could also be run in the other direction, using quantitative methods as a means of strengthening confidence in the evidence already gleaned from the qualitative methods.



Note: The three-dimensional graph shows the slope expressing the relation between income and life expectancy. The size of the bubbles represents population. *Gapminder* also permits analyses of other variables including economic growth, income inequality, health, education, the environment, and other measures. Analyses may be projected over 50-year periods and longer.

Source: *Gapminder*: <https://www.gapminder.org/>.

Figure 13.5 Analysis of life expectancy and income in 2018

The use of Bayesian methods has one major disadvantage compared with the other methods mentioned above. The disadvantage is that it depends on having a clear hypothesis and having agreed upon measures of the phenomena being investigated. Both nested analysis and the informal comparison of cases allow for more exploratory investigations of relationships and may encourage theory building rather than theory testing. While Bayesian analysis is implicit in process-tracing as a qualitative method (Rohlfing 2012), when used to integrate within-case and cross-case findings some of the inductive capacity of this integration of methods can be diminished or lost entirely.

The above said, using the Bayesian logic does provide a more clearly defined principle for integrating findings from alternative research styles than do the other methods we have been discussing. For example, nested analysis does discuss the interaction of quantitative and qualitative results as a goal of the analysis, but relies largely on the judgment of the researcher to make the integration, and to decide when there have been sufficient iterations between quantitative and qualitative analyses to say that a satisfactory solution has been obtained. While we would not want to diminish the importance of good judgment by researchers, we also recognize that having additional standards for making the judgments can be beneficial (see Gordon and Smith 2004; and Figure 14.4).

3.8 Qualitative Case Analysis Design

Qualitative Comparative Analysis (QCA) represents one of the major alternatives to the informal approach to combining multiple cases, and making comparative statements based on case study research (Ragin 2008; Thomann, Chapter 15 this volume). The logic of QCA is to identify necessary and sufficient conditions for the occurrence of a particular outcome of a case – the analogue of a dependent variable.

There are, of course, significant dangers involved in the use of existing cases, whether done informally or done more formally through QCA. The most obvious is that the researchers involved may be the main independent variable. That is, when individuals are conducting case research for their own theoretical and empirical purposes, they will inevitably bring their own biases and predispositions into the research. Those biases may be apparent in the research, or they may not, so that the individual utilizing multiple cases is potentially working with cases that have unknown sources of extraneous variance hidden within the findings (see Peters 2013).

4. BOUNDING MULTIPLE TRIANGULATION

It is clearly unfeasible to employ more than a few of the above designs in the same project. For this reason, the designs are presented as opportunities for realizing some of the benefits arrayed in Figure 13.1. Although now commonly denigrated within the social sciences, the case method of research still provides the researcher with a great deal of useful information. The principal advantage of the case method is that it provides rich insights into the dynamics of decision making, and further places that decision making into a context (see Pollitt 2013). With a case, the researcher can determine better than with other, more popular social science methods, what was actually happening to produce a public policy, or to make a policy work through the implementation process. This utility for the case method is true whether one

utilizes more traditional forms of case analysis (see Haverland and Blatter 2012) or utilizes process-tracing (see Beach, Chapter 14 this volume).

The absence of the capacity to generalize has become something of a cliché and is taken as a reason by some scholars to dismiss case research. That dismissal is generally too facile and fails to recognize the importance of intensive case research for understanding causal mechanisms and causal processes and even generalizing findings based on the principle of proximal similarity, rather than with the dubious method of calculating p-values (Campbell 1975). This capacity to identify causal processes and the existence of causal triggers working within a particular context is central to case-based methodologies, especially process-tracing (Beach 2016; Beach and Pedersen 2013).

As we have seen, we can build on the availability of multiple cases, whether the cases are conducted purposefully to investigate a phenomenon in different settings, or they have been done previously for other purposes and can be identified as being on the same subject. The library shelves are filled with cases, as are sites such as the Development Experience Clearinghouse (DEC) and these can be marshaled to do comparative analysis. That presents the opportunity for scholars to build cross-case analysis on these multiple available studies that focus on within-case analysis.

There are at least two ways in which scholars can combine available, or purpose-built cases, in order to engage in meaningful cross-case analysis. One is a relatively informal method in which the analyst simply examines the results of several cases and attempts to draw conclusions based on the evidence in each. One of the clearest examples of this method, in particular for public policy, is Hugh Hecló's (1974) study of social policy in Britain and Sweden. Peter Hall's research comparing economic policy in Britain and France (1986) would also fall into this category of cross-case analysis.

While this informal method of comparing cases appears wildly "unscientific" having no clear methodological guidelines and no established way of assessing the degree of similarity or difference across cases, it also has its uses. The individual researcher may be more capable of dealing with the inconsistencies among case purposes and internal logics than can a formal method for doing so. The reliance on the individual researcher may introduce another source of invalidity into the research, but perhaps may produce fewer false positives or false negatives than would a more formal analysis. This reliance on human judgment is especially important for cases that are more found than produced.

In addition to biases created by multiple investigators, who have different purposes for including cases in their sample, there may be other sources of invalidity in utilizing multiple cases. For example, studies coded at different times may encounter the challenges of history (see Campbell and Stanley 1963) and other threats to validity that affect qualitative as well as quantitative research (Shadish, Cook, and Campbell 2002). Researchers in different political systems may have different levels of access to information and to interviews that might be available in other settings.

5. CONCLUSIONS

The above discussion has approached within-case and cross-case analysis in accordance with a general model of mixed-methods research (Figure 13.1). The general model is grounded in what we see as an evolving theory of critical multiplism and mixed-methods triangulation.

A range of alternative multi-method designs build on the general model and the theory of critical realism which supports it. We have also considered special issues that arise when analyzing public policy across political systems. Some of these points have been intimated above and we have attempted to develop them in more detail. In so far as possible we have attempted to illustrate the points being made using the available literature in comparative policy, complemented by sources that are more multi-disciplinary.

Perhaps the most obvious point we have made about public policy as opposed to other areas of comparative analysis is the importance of context. Any scholar of comparative politics will immediately acknowledge that context is important for their research.⁸ The continuing diffusion of policy ideas and practices represents another important reason for integrating within-case and cross-case analysis. Public policy more than other areas of the social sciences involves the purposive search for lessons from other policy making domains. As we know from Galton's problem, different social systems may have similar attributes, but it can be difficult to ascertain if these represent autonomous choices or the results of diffusion. Especially with the contemporary importance of "evidence-based policy making" (Sherman 2003) perhaps the only way to sort out the apparent similarities in policy adoption is using within-case analysis.

Public policies are purposefully designed. Rather than resulting from a random confluence of variables, they are the consequence of the intersection of causal agency, provided by practical reasoning, and robust causal mechanisms. Cross-case analysis, because it deals with surface characteristics of policies, provides limited information about the causal agency and mechanisms that drive the design and implementation of policies. Instead, by using mixed methods to inquire into what occurs within as well as across cases, we may take the study of comparative politics and public policy further along the road to a more robust understanding of the field.

NOTES

1. Triangulation raises questions about alternative theories of validity or "truth". Generally, there are three theories: correspondence theory, coherence theory, and pragmatic theory (Shadish, Cook, and Campbell 2002: 35–6). The three theories are frequently used concurrently, although only positivism, at least in its early form, employs the correspondence theory exclusively. Because the coherence theory requires correspondence in order to avoid relativism, the pragmatists coined the term objective relativism (see Dunn 2019). Perhaps a more precise term is William Whewell's consilience of inductions. Whewell (1794–1866) held that a consilience of inductions is a test of the truth (validity) of a theory. Consilience occurs when an induction based on one class of facts coincides with an induction based on another. See Laudan (1971) and Wilson (1999).
2. Traditionally, the concept of complementarity has arisen only in connection with the joint use of qualitative and quantitative methods.
3. Some 1,257 were case studies, most of which were unstructured and could not be coded.
4. In meta-regression analysis, samples of original meta-analyses are the units of analysis. Effect sizes, the dependent variable, are typically regressed on independent variables including contents of civic education training programs, length of programs, age of participants, and so forth. A binary measure such as political participation or its absence might also constitute the dependent variable, indicating among other things what type of causal model was used.
5. Campbell also recommends the replacement of internal validity with local molar causal validity in order to denote the recommended substitution of micro-causal relations ("molecules") with macro-causal ("molar") relations.
6. Another design might employ canonical correlation or multiple regression analysis with dummy variables (presence or absence of the policy and type of climate).

7. In the following extract the term “ethnographer” has been replaced with “researcher”. This does not change Campbell’s analysis and conclusions.
8. Remember, however, that Przeworski and Teune (1970) had argued for the elimination of country names, and context, in comparative analysis so that the research could proceed using only variable names.

REFERENCES

- Abel, T. (1948). The operation called Verstehen. *American Journal of Sociology* 54 (3): 211–18.
- Abel, T. (1975). Verstehen I and Verstehen II. *Theory and Decision* 6 (1): 99–102.
- Adcock, R. and Collier, D. (2001). Measurement validity: a shared standard for qualitative and quantitative research. *American Political Science Review* 95: 529–46.
- Ahmed, A. and Sil, R. (2012). When multi-method research subverts methodological pluralism – or, why we still need single-method research. *Perspectives on Politics* 10 (4): 935–53.
- Angrist, J. D. and Pischke, J. S. (2008). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princeton University Press.
- Angrist, J. D. and Pischke, J. S. (2014). *Mastering Metrics: The Path from Cause to Effect*. Princeton, NJ: Princeton University Press.
- Beach, D. (2016). It’s all about mechanisms: what process tracing case studies should be tracing. *New Political Economy* 21: 463–72.
- Beach, D. and Pedersen, R. B. (2013). *Process Tracing Methods: Foundations and Guidelines*. Ann Arbor, MI: University of Michigan Press.
- Beck, N. (2010). Causal process “observation”: oxymoron or (fine) old wine? *Political Analysis* 18 (4): 499–505.
- Bellino, M. J. (2015). Civic engagement in extreme times: the remaking of justice among Guatemala’s “postwar” generation. *Education, Citizenship and Social Justice* 10 (2): 118–32.
- Bennett, A. (2008). Process tracing: a Bayesian perspective. In J. M. Box-Steffensmeier, H. E. Brady, and D. Collier (Eds.), *The Oxford Handbook of Political Methodology*. Oxford: Oxford University Press, 702–21.
- Bhaskar, R. (1975). *A Realist Theory of Science*. Leeds: Leeds Books.
- Bhaskar, R. (1979). *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*. Brighton: Harvester Press.
- Brown, S. R. (1993). A primer on Q methodology. *Operant Subjectivity* 16 (3–4): 91–138.
- Brunner, R. (1986). Case-wise policy information systems: redefining poverty. *Policy Sciences* 19 (2): 201–23.
- Campbell, D. T. (1975). “Degrees of freedom” and the case study. *Comparative Political Studies* 8 (2): 178–93.
- Campbell, D. T. (1986). Relabeling internal and external validity for applied social scientists. *New Directions for Evaluation, Special Issue: Advances in Quasi-Experimental Design and Analysis* 31: 67–77.
- Campbell, D. T. and Stanley, J. C. (1963). *Experimental and Quasi-Experimental Design for Research*. Chicago, IL: Rand-McNally.
- Collingwood, R. G. (1940). *An Essay on Metaphysics*. Oxford: Clarendon Press.
- Cook, T. D. (1985). Postpositivist critical multiplism. In L. Shotland and M. M. Mark (Eds.), *Social Science and Social Policy*. Beverly Hills, CA: Sage, 21–62.
- Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.
- Dunn, W. N. (2019). *Pragmatism and the Origin of the Policy Sciences: Harold Lasswell and the Chicago School*. Cambridge: Cambridge University Press.
- Dunn, W. N. and Swierczek, F. W. (1977). Planned organizational change: toward grounded theory. *Journal of Applied Behavioral Science* 13 (2): 135–57.
- Fisher, R. (1949). *The Design of Experiments*. Edinburgh: Oliver and Boyd.
- Flick, U. (2007). *Qualitative Research Designs: Designing Qualitative Research*. London: Sage.

- Gasking, D. (1955). Causation and recipes. *Mind* 64: 479–87.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Boca Raton, FL: CRC Press.
- Glymour, C. N. (2001). *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*. Cambridge, MA: MIT Press.
- Gordon, S. C. and Smith, A. (2004). Quantitative leverage through qualitative knowledge: augmenting the statistical analysis of complex causes. *Political Analysis* 12: 233–55.
- Hall, P. A. (1986). *Governing the Economy: The Politics of State Intervention in Britain and France*. Oxford: Oxford University Press.
- Haverland, M. and Blatter, J. (2012). *Designing Case Studies: Explanatory Approaches in Small-N Research*. Basingstoke: Palgrave Macmillan.
- Heclo, H. (1974). *Modern Social Politics in Britain and Sweden*. New Haven, CT: Yale University Press.
- Hirst, E., Goeltz, R., and Trumble, D. (1987). *Electricity Use and Savings in the Hood River Conservation Project* (No. ORNL/CON-231). Oak Ridge National Lab., TN (USA).
- Horwich, P. (1990). *Truth*. Oxford: Clarendon Press.
- Houts, A. C., Cook, T. D., and Shadish Jr, W. R. (1986). The person–situation debate: a critical multiplist perspective. *Journal of Personality* 54 (1): 52–105.
- Humphreys, M. and Jacobs, A. (2015). Mixing methods: a Bayesian approach. *American Political Science Review* 109: 653–73.
- Johnson, R. B., Onwuegbuzie, A. J., and Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research* 1 (2): 112–33.
- Kasza, G. J. (2005). Quantitative methods: reflections on the files of recent job applicants. In K. R. Monroe (Ed.), *Perestroika: The Raucous Rebellion in Political Science*. New Haven, CT: Yale University Press, 421–33.
- Kempton, W. (1986). Two theories of home heat control. *Cognitive Science* 10: 75–90.
- Kennedy, P. (2003). *A Guide to Econometrics*. Cambridge, MA: MIT Press.
- King, G., Keohane, R. O., and Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Kirk, R. E. (1982). *Experimental Design: Procedures for the Behavioral Sciences* (2nd edition). New York: Brooks Cole.
- Laudan, L. (1971). William Whewell on the consilience of inductions. *The Monist* 55 (3): 368–91.
- Li, Q., Owen, E., and Mitchell, A. (2018). Why do democracies attract more or less foreign direct investment? A meta-regression analysis. *International Studies Quarterly* 62 (3): 494–504.
- Lieberman, E. S. (2005). Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review* 99: 435–52.
- Lijphart, A. (1971). Comparative politics and the comparative method. *American Political Science Review* 65: 682–93.
- Mackie, J. L. (1984). *The Cement of the Universe: A Study of Causation*. Oxford: Oxford University Press.
- Mahoney, J. (2007). Qualitative methodology and comparative politics. *Comparative Political Studies* 40 (2): 122–44.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review* 62: 279–300.
- Miles, M. B. and Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA: Sage.
- Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.
- Peach, H. G., Peters, D., Oliver, T., and Goldstein, D. (1984). Cooperation and diversity in a large-scale conservation research project. In *ACEEE Summer Study on Energy Efficiency in Buildings*, vol. I. Washington, DC: American Council for an Energy-Efficient Economy, 112–35.
- Pearl, J., Glymour, M., and Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. Chichester: John Wiley & Sons.
- Pearl, J. and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.

- Peters, B. G. (2013). *Strategies for Comparative Research in Comparative Politics*. Basingstoke: Palgrave Macmillan.
- Pollitt, C. (2013). *Context in Public Policy and Management: The Missing Link?* Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Przeworski, A. and Teune, H. (1970). *The Logic of Comparative Social Inquiry*. New York: John Wiley.
- Ragin, C. (2008). *Redesigning Social Inquiry: Set Relationships in Social Research*. Chicago, IL: University of Chicago Press.
- Rohlfing, I. (2012). *Case Studies and Causal Inference: An Integrative Framework*. New York: Palgrave Macmillan.
- Shadish, W. R. and Cook, T. D. (1999). Comment-design rules: more steps toward a complete theory of quasi-experimentation. *Statistical Science* 14 (3): 294–300.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Shadish, W. R., Cook, T. D., and Houts, A. C. (1986). Quasi-experimentation in a critical multiplism mode. *New Directions for Program Evaluation* 31: 29–46.
- Sherman, L. W. (2003). Misleading evidence and evidence-led policy: making the social sciences more experimental. *Annals of the American Academy of Political and Social Science* 589: 6–19.
- Skocpol, T. (1979). *States and Social Revolutions: A Comparative Analysis of France, Russia and China*. Cambridge: Cambridge University Press.
- Thomas, M. A. (2010). What do the worldwide governance indicators measure? *European Journal of Development Research* 22 (1): 31–54.
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.
- Udy Jr, S. H. (1959). “Bureaucracy” and “rationality” in Weber’s organization theory: an empirical study. *American Sociological Review* 24 (6): 791–5.
- von Wright, G. H. (1971). *Explanation and Understanding*. Ithaca, NY: Cornell University Press.
- Wallace, W. (1971). *The Logic of Science in Sociology*. Chicago, IL: Aldine.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. (1999). *Unobtrusive Measures*. Newbury Park, CA: Sage.
- Weber, M. (1946). *From Max Weber: Essays in Sociology*. New York: Oxford University Press.
- Wilson, E. O. (1999). *Consilience: The Unity of Knowledge*. New York: Vintage.
- Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- Yin, R. K. (2017). *Case Study Research and Applications: Design and Methods*. Thousand Oaks, CA: Sage.
- Yin, R. K. and Heald, K. A. (1975). Using the case survey method to analyze policy studies. *Administrative Science Quarterly* 20 (3): 371–81.
- Ziliak, S. and McCloskey, D. N. (2008). *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor, MI: University of Michigan Press.

14. Causal case studies for comparative policy analysis

Derek Beach

1. INTRODUCTION

Causal case studies involve the in-depth assessment of how policy processes play out within cases. Given that the core analysis is within-case, a core challenge in policy studies is assessing whether similar processes are operative in multiple cases. At first glance, comparative analysis using tools like a most-similar systems design (MSSD) might seem to be the method of choice to accomplish this. However, MSSD comparisons only enable the assessment of similarities and differences in causes/contextual factors and outcomes across cases – in effect assessing input and output. By their very nature, MSSDs are variance-based, *cross-case* comparisons that shed no light on the policy processes that are actually operative *within* cases, linking causes and outcomes together. This is because a MSSD is basically a quasi-experiment using observational data, and despite all of their virtues, experiments tell us nothing about *how* an effect is produced *within* a case (Dowe 2011; Illari 2011; Machamer 2004; Russo and Williamson 2007).

In this chapter I argue that a case-based approach is the only viable design for studying *how policy processes actually play out in real-world cases*. In a case-based approach, the analytical priority is the in-depth study of operative processes within a case. Comparisons act as an adjunct method used to map populations into smaller, relatively homogeneous sub-sets that enable us to select relevant cases for within-case analysis. The alternative to using detailed within-case studies is to attempt to use variance-based comparisons that end up black-boxing the processes at play because they involve assessing the *difference* that *variation* in values of a cause make for values of the outcome *across* cases using different forms of comparisons.

The key strength of case-based approaches to public policy is that we learn how a causal process actually works in a given case, or bounded set of cases in a similar context. However, one of the downsides is that we are left in the dark regarding how a process works within a larger, more diverse population. The crux of the challenge is that studying how policy processes play out within cases requires that we significantly lower the level of abstraction of our analysis. However, this means that our findings about how processes play out in particular cases become highly sensitive to contextual differences across cases, meaning that it can be difficult to make meaningful generalizations about how policy processes actually work that hold for cases in different contexts. And even when two cases look similar at the cross-case level (similar values of the cause(s) and contextual factors and outcome), this does not mean that similar policy processes were operative within the cases because there can be unknown conditions that impact how a process works. In essence, in case-based approaches to studying public policy, we learn a lot about a little.

The chapter proceeds in three steps. After defining core terms like case and causal, the chapter uncovers the fundamental ontological and epistemological assumptions that separate

case-based from variance-based approaches to studying public policy. The chapter then explores how case-based methods that combine the comparative mapping of a set of cases aimed at creating as homogeneous a set of cases as possible, and the in-depth tracing of mechanisms (i.e. processes) within multiple cases in this bounded population enable valid causal inferences to be made about policy *processes* operative *within* cases that share a set of contextual conditions (aka scope conditions).

2. KEY TERMS

A *case* is an instance of a causal theory playing out from start to finish. Irrespective of approach, the level at which causes are operative is always within a single case. An increase in the number of veto players can produce deadlock and joint decision-traps within a political system, but a change in one country would not produce deadlock *across* different countries unless there are diffusion or other dependencies across cases. One can *learn* about the effect that the increase in veto players has using variance-based methods by comparing a case where this took place with one where it was absent, where all other things are equal. This would tell us whether variation in the cause resulted in variation in the outcome, controlled for potential confounders. However, while we might study the causation by comparing across cases, at the end of the day causation always occurs *within* cases.

Where case-based and variance-based approaches differ is in their understanding of the term *causal* (ontological level), and what type of empirical evidence can be used to make valid *causal inferences* (epistemological level). At the ontological level, the core distinction is whether causation is understood in *counterfactual* terms (variance-based) (Woodward 2003), or in *mechanistic* terms (case-based) (Illari and Williamson 2011; Machamer, Darden, and Craver 2000; Waskan 2011). Counterfactual causation is defined as the claim that a cause (or mechanism) produced an outcome because its absence would result in the absence of the outcome, all other things being held equal (Woodward 2003). The essence of mechanistic causal claims is that we shift the analytical focus from identifying causal effects to explaining how causes are linked to an outcome – i.e. from causes → outcomes to the process *in-between*. Mechanisms are *not* causes, but are causal processes that are *triggered* by causes and that *link* them with outcomes in a productive relationship (Bunge 1997; Glennan 1996; Machamer 2004; Machamer, Craver, and Darden 2000; Russo and Williamson 2007, 2011). The epistemological distinction that flows from this ontological difference relates to how one learns about causal relationships. Learning about counterfactual causal relationships requires assessing the difference that variation in the cause makes for values of the outcome across cases, whereas case-based approaches study causation by tracing the manifestations left by the operation of mechanisms within cases (i.e. mechanistic evidence).

Finally, we need to specify the meaning of the term *causal heterogeneity*, because it can refer to both situations where causes work differently, or where different mechanisms are operative. Causal heterogeneity refers here to all types of causal complexity at the level of causes/outcomes across a set of cases, including situations where the same cause can produce different outcomes in different contexts (multifinality), different causes can produce the same outcome in different contexts (equifinality), and where the nature of a relationship differs across cases (e.g. positive in cases where factor Z1 is present, negative when factor Z1 is absent). In contrast, the term mechanistic heterogeneity is reserved for the situation where the

same cause and outcome are linked together through different mechanisms in different contexts, or the same cause triggers different mechanisms that are linked to different outcomes. We now turn to a discussion of the two approaches to studying policy processes.

3. VARIANCE-BASED VERSUS CASE-BASED APPROACHES TO STUDYING PUBLIC POLICIES

Based on recent developments in the philosophy of science and social science methodology, we can distinguish between variance-based (“top-down”) and case-based (“bottom-up”) approaches to public policy research (Beach and Pedersen 2016; Cartwright 2011; Goertz and Mahoney 2012; Ragin 2000; Russo and Williamson 2011). The variance-based approach typically uses large-N comparisons (i.e. statistical methods), although case studies in the form of different types of focused cross-case comparisons can also be used (e.g. Gerring 2017; King, Keohane, and Verba 1994). The term case-based approach is sometimes also termed “qualitative”, although this usage is imprecise because qualitative is also used to refer to a variety of interpretivist methods. The core of case-based methods is within-case tracing of causal mechanisms using process tracing, although cross-case comparisons are important for selecting appropriate cases and bounding populations into relatively causally homogeneous sets (see Schneider and Rohlfing 2013, 2016).

3.1 Top-Down, Variance-Based Approaches: “It Works Somewhere” Claims

Nancy Cartwright has succinctly defined the essence of the types of claims about mean causal effects that variance-based approaches enable; “it works somewhere” (2011). In variance-based approaches, the methodological gold-standard is an actual experiment (randomized controlled trial – RCT), which if properly designed, enables strong causal inferences about the *mean causal effect* of a given treatment variable within the studied sample (Clarke et al. 2014; Gerring 2011).

Variance-based approaches build on a *counterfactual* understanding of causation – often developed as the potential outcomes framework (Angrist and Pischke 2009; Rubin 2005; Woodward 2003). Without evaluating the difference that a cause can make between the actual and the counterfactual, no causal inferences about counterfactuals are possible. To assess a counterfactual causal claim, one therefore needs to empirically assess the counterfactual (aka the potential outcome), holding the impact of all other potential causes and confounders constant (i.e. a MSSD). This is relatively easy to see in an experiment, where we compare values of the outcome in cases that receive the treatment with those in the experimental control group that do not (i.e. the counterfactual state). Here the lack of treatment in the control group acts as the counterfactual, enabling us to infer that if there is a significant and substantial difference in values of the outcome in the two groups, this difference is the mean causal effect of the treatment. Given the need to compare *across* cases, variance-based approaches can be termed a “top-down” form of research (Illari and Williamson 2011). Again, this is best seen in an experiment, where mean causal effects (the average “difference” that the cause makes for the outcome across the treatment and control groups) are assessed within the population of cases in the study. The term top-down is therefore appropriate because causation is studied at the population-level (or samples thereof) by assessing trends *across* cases.

Strictly speaking, observational data in the form of statistical co-variation of causes and outcomes across many cases does not enable causal inferences to be made unless we assume that the data has the character of a natural experiment that enables us to claim that our population is split (either temporally or spatially) into a treatment and control group in which everything else is constant (Angrist and Pischke 2009). Even more challenging is the claim that we can make causal claims based on counterfactuals when studying single cases. One way of proceeding is to transform “one case into many” by disaggregating a case either spatially or temporally, enabling an assessment of the counterfactual in the form of a most-similar system comparison (everything else is equal except variation in the cause) (King, Keohane, and Verba 1994: 217–28). Another way of doing variance-based case studies involves using counterfactual single case studies, where hypothetical evidence about “what might have been” is used as the counterfactual comparison. Single-case, counterfactual comparisons therefore involve comparing an existing real-world case with a hypothetical counterfactual case, where the logical argument is then made that if a particular cause had not occurred, the outcome would not have occurred (Goertz and Levy 2007; Lebow 2000; Levy 2015; Tetlock and Belkin 1996).

Key to the ability to make inferences about mean causal effects are the assumptions of unit homogeneity and independence of units (Holland 1986; King, Keohane, and Verba 1994: 91–7). *Unit homogeneity* means that the same cause will produce the same results in two or more cases (i.e. causal homogeneity, also termed stable unit treatment effect, see Morgan and Winship 2007: 37–40). *Independence* means that the potential outcome in one case is unaffected by values of the cause in other cases. If these two assumptions do not hold, we will have biased estimates of the difference that variations in X have for values of Y.

In variance-based research, these two assumptions hold when we have *many units* that are *randomly selected* into treatment and control groups, and where everything else is constant, thereby ensuring that any differences in how a treatment works between units washes out at the level of comparisons of large groups. Independence is ensured in an experiment through random selection, where the values of X are independent of values taken by Y.

In variance-based approaches there is a clear *evidential hierarchy* that relates to the strength of causal inferences enabled within the given study (i.e. *internal validity*) with respect to whether these two assumptions hold (Clarke et al. 2014; Gerring 2011). Actual experimental designs are at the top,¹ enabling strong causal inferences to be made, followed by natural experiments using observational data where one can assume that the treatment and control were “randomly” assigned by nature.

Findings from case studies are at the bottom of the evidential hierarchy because they tell us precious little about *trends* when causal heterogeneity is present in a population. Unfortunately, the assumptions of unit homogeneity and independence almost never hold when engaging in a small-N comparison of difference-making. For example, almost any one-into-many transformation of cases will result in a set of cases that are not causally similar, and there will also be serious violations of case independence where values of X in one case will be affected by values of Y in preceding or simultaneously occurring cases. With regard to unit homogeneity, disaggregating a negotiation as a case temporally into stages (t_0, t_1, \dots, t_n) results in cases that are quite causally dissimilar, where we can expect critical differences in how causes/mechanisms play out when comparing early stages (agenda-setting) and the end game. In addition, the “cases” would not be independent of each other, because in a policy process, what happens at the start (t_0) naturally affects events later in the negotiation, meaning that values of

Y in case t_0 will influence values of X in subsequent cases (periods of the negotiation). If we disaggregated a policy negotiation into different issue areas instead of temporally, we should expect that deals or deadlock with respect to one issue (case) will affect other important issues (other cases), especially in a setting where package deals are typical forms of resolving negotiations. The different “cases” would also not be homogeneous in that we would expect that factors such as expertise might matter more in low-salience issues and matter less in highly salient issues in which actors would have incentives to mobilize the necessary informational resources to understand an issue.

Similar problems occur when we try to identify two or more cases that can be compared in a MSSD. As Runhardt (2015: 1306) admits, “A similarity comparison in areas like political science is, however, difficult to defend.” Because of the complexity of the social world, it is difficult to find cases in which the “all other things equal” assumption required in a natural experiment actually holds (Ragin 1987: 48). Levy (2015: 390) writes that “Controlled comparison and matching face the difficulty of finding real-world cases that are identical in all respects but one.” But unless we can substantiate that all other things are equal except for the presence/absence of a cause, we cannot make a causal inference that its absence made a difference for the outcome.

Therefore, at best, case studies can help us detect measurement error or find potential confounders when engaging in more exploratory research that can help us improve the statistical models we use to explore population-wide difference-making (Seawright 2016: 45–69).

There are two critical weaknesses of top-down, variance-based research that make it difficult to communicate meaningfully with case-based research; one of which can be resolved to some extent, the other not. First, variance-based approaches shed light on mean causal effects across a set of cases. If a population was completely causally homogeneous, the unit homogeneity (stable unit treatment effect) would hold perfectly (Morgan and Winship 2007: 37–40), meaning that population-level trends would be perfectly predictive for effects in individual cases (Cartwright 1999: 154–9). But given the causal complexity of the real world, there can be many reasons that the average causal effect across a study population does not perfectly correspond to the local causal effect within a case (Cartwright 2012: 980–981; Leamer 2010; Williams and Dyer 2009: 85–7). Because of this, *ontologically probabilistic* claims are made about trends (i.e. mean causal effects) in variance-based approaches.

When one then moves from population-level causal claims about trends to individual cases, causes therefore become “probability-raisers” (Gerring 2011: 199). If there is a positive relationship between X and Y, a high value of X would make it more probable that we would find a high value of Y in case A. Based on what we know about mean causal effects of different independent variables and the impact of confounders, propensity scores can in theory then be estimated for individual cases. However, to do this requires either that we have evidence of a high level of causal homogeneity in the population being studied that enables one to assume overall treatment effects apply to individual cases in a predictable fashion, or we have in effect mapped the causal heterogeneity embedded within the population, enabling cases to be grouped together into more homogeneous sub-sets of cases (e.g. there is a negative relationship between X and Y when factor Z1 is present, whereas there is no relationship in cases where factor Z1 is absent) (Leamer 2010). If neither holds, there is the significant risk of an *ecological fallacy* when inferring from population-level trends to individual cases (Robinson 1950). Actual experiments have the further difficulty that their inferences do not necessarily

hold outside the controlled laboratory setting, meaning that the ability to infer to cases outside the lab is even further reduced.

Second, even if we would be able to accurately estimate propensity scores for individual cases, studying causal claims by comparing values of X and Y *across* cases does not tell us how policy processes work *within* a case. In other words, we learn about the difference variation in X makes for values of Y, but we do not learn anything about the causal arrow linking the two – it remains firmly within a black box. An experiment does not tell us *how* a treatment works – only that there is a mean causal effect (Dowe 2011; Illari 2011; Machamer 2004; Russo and Williamson 2007). In order to learn about how causes actually work within cases, we need to move away from counterfactual difference-making to explore how policy processes play out in actual cases.

Concluding, variance-based approaches are top-down methods that assess counterfactual causation in the form of mean causal effects across cases. The relative strengths of the approach are the ability to assess the magnitude of net causal effects, and our ability to make causal inferences about many cases (populations or samples thereof). The core weakness relates to our ability to say anything meaningful about individual cases because of the risk of causal heterogeneity within populations, meaning that at most we can make educated guesses using case propensity scores.

3.2 Bottom-Up Case-Based Approaches: How Policy Processes Work in Cases

Case-based approaches can be understood as being “bottom-up” designs because the in-depth study of individual cases is the analytical point of departure. Here the goal is to learn about causal mechanisms and how they operate in particular cases (Russo and Williamson 2011). Mechanisms are not causes; they are what link causes and outcomes together. In research on public policy, they are the policy processes that link different causes with outcomes within cases.

In a case-based understanding, causal mechanisms are more than just lower-level counterfactual claims. Mechanisms are viewed in a more holistic fashion than mere counterfactuals; meaning the effects of a mechanism are more than the sum of their parts. If one takes mechanisms seriously, the goal is to explore what process *actually* was operative in a case (Groff 2011; Machamer 2004: 31; Waskan 2011). A “mechanism explanation for some happening that perplexes us is explanatory precisely in virtue of its capacity to enable us to understand how the parts of some system actually conspire to produce that happening” (Waskan 2011: 393). The essence of mechanistic explanations is that we shift the analytical focus from causes and outcomes to the hypothesized causal process in between them. That is, mechanisms are not causes but are causal processes that are *triggered* by causes and that link them with outcomes in a productive relationship.

In case-based approaches, the focus is on tracing the operation of causal mechanisms within cases (Beach and Pedersen 2016, 2019). The key elements of a causal mechanism are unpacked theoretically and studied empirically in the form of the traces left by the activities associated with each part of the policy process. Each of the parts of the mechanism can be described in terms of entities that engage in activities (Machamer 2004; Machamer, Darden, and Craver 2000). Entities are the factors (actors, organizations, or structures) engaging in activities, whereas the activities are the producers of change or what transmits causal forces or powers through a mechanism. When a causal mechanism is unpacked theoretically as a system, the

goal becomes understanding how a process actually works by tracing the operation of each part (or at least the most critical parts) in one or more cases.

Mechanisms are traced empirically by collecting *mechanistic evidence*, which are the observable fingerprints left by the operation of the activities associated with parts of mechanisms (Illari 2011; Russo and Williamson 2007). Here there is *no variation*; instead it is the empirical traces and their association with activities that enable us to infer that we have evidence of a mechanism linking a cause (or set of causes) with an outcome (Beach and Pedersen 2019; Clarke et al. 2014). Mechanistic evidence is observational data, trying to capture what really took place within individual cases.

In case-based research, detailed tracing of policy processes using mechanistic evidence within individual cases is at the top of the evidential hierarchy. Below this are weaker within-case methods that only obliquely trace mechanisms (congruence studies and analytical narratives), thereby not enabling strong causal inferences. At the bottom are comparisons across cases using methods like QCA that can be used to detect potential causes, select appropriate cases for within-case analysis, and enable cautious generalizations about processes to small, bounded sets of cases.

There are two weaknesses of case-based approaches that are in many respects the antithesis of variance-based approaches. First, taking individual cases as an analytical point of departure requires making *deterministic* causal claims about mechanisms (Beach and Pedersen 2016: 19–24; Mahoney 2008). Ontological determinism can be defined simply as the claim that in a given case, an outcome has occurred for a reason (or more realistically, a set of reasons) (Adcock 2007; Bhaskar 1979: 70–71; Mahoney 2008). This is a form of bottom-up claim about causation (Russo and Williamson 2011), in which outcomes have occurred for a set of reasons within a particular case. But this then creates the challenge that moving upwards from the individual case to other cases can only realistically be done when we translate our claims into probabilistic ones except if the target population of the inference exhibits a “law-like” causal homogeneity (Cartwright 1999: 154–9), meaning that the same cause produces the same outcome through the same process in *all of the cases* in the population. This does not mean that we empirically are always able to figure out why something happened in a case (epistemological level). However, just because we cannot figure out empirically why something occurred does not mean that the outcome was the product of randomness. Things do not “just happen” in cases; things happen for a reason.

The implication of this is that detailed knowledge about how a policy process is operative in a single case cannot easily be exported to other cases because mechanisms at this level of detail are sensitive to even relatively slight contextual differences (Bunge 1997; Falleti and Lynch 2009; Gerring 2010; Goertz and Mahoney 2009). In Cartwright’s language, we learn about how “it works here”, but it is difficult to extrapolate that it also “works there” (2012). This means that mechanistic heterogeneity can be produced by contextual differences, defined as situations: (1) where the same causes trigger different processes in two or more cases, thereby resulting in different outcomes; or (2) where the same cause is linked to the same outcome through different processes. The risk of the first variant can be reduced through careful mapping of the population by scoring cases on their values of the cause, outcome, and relevant contextual conditions. However, the second scenario is more problematic because mechanistic heterogeneity might be lurking under what might look like a homogeneous set of cases at the level of causes/outcomes. Given this sensitivity, our ability to generalize from studied cases to other cases using comparisons is significantly weakened. We trade higher internal validity of

causal inferences about policy processes for more limited ability to generalize outside of the studied population (i.e. lower external validity). Extrapolating from the individual (or small group) to the full population in this situation would result in an *atomist fallacy*.

A real-world example of mechanistic heterogeneity produced by differences in context can be found in White (2009). He describes a mechanism that links a policy intervention (cause = education of mothers in nutrition) with an outcome (improved nutritional outcomes for children) that was found to have worked in a case (the Tamil Nadu Integrated Nutrition Project in India). The unpacked mechanism can be described as: Cause (mother participates in program) → (1) mother receives nutritional counseling → (2) exposure results in knowledge acquisition → (3) knowledge used to change child nutrition → Outcome (improved nutritional outcomes) (based on White 2009: 4–5). Based on the success of the program in the Tamil Nadu case in India, it was then attempted to use the same policy intervention in Bangladesh. However, the process did not function as expected in the different context; instead it broke down. The reason for this was a key contextual difference. In Bangladesh, mothers were not the key decision-makers in households, with men doing the shopping, and mothers-in-laws in joint households (sizeable minority) acting as decision-makers about what food went onto the table. The mechanism therefore “worked” until step 3, but because of a contextual difference, it broke down in the Bangladesh case.

The alternative to taking mechanistic heterogeneity seriously by appreciating the complexity of real-world cases and the limited bounds of generalization of mechanisms because of contextual sensitivity is to lift the level of abstraction about theorized policy processes to such a high level that our theorized mechanisms are in essence “nothing-burgers” that tell us precious little, if anything, about how a process works in real-world cases. Therefore, case-based research only attempts to generalize to relatively small, bounded populations of cases that are relatively homogeneous (both at the level of causes and mechanisms).

4. HOW TO DO CASE-BASED STUDIES OF POLICY PROCESSES

After having detailed the core differences between variance-based and case-based approaches to research, the chapter turns to an exposition of what a cumulative, case-based approach that combines intensive within-case analysis and process-focused comparisons to studying public policy looks like.

The first step in a case-based research design focused on studying policy processes is to identify the cause(s) and outcome of interest.² After defining key concepts such as policy crisis (causes, outcomes, and potentially relevant contextual conditions) in a set-theoretic manner (also termed crisp-sets), where the critical distinction becomes whether a case is a member of a given concept or not,³ a determination needs to be made about whether it is possible to theorize a plausible mechanism linking the cause(s) and outcome, or whether a theory-building approach needs to be chosen. If it is possible to theorize a plausible policy process, it should be unpacked into parts composed of entities engaging in activities. Activities are the causal link that binds the parts of the mechanism together, ideally into an unbroken chain between the cause(s) that trigger the process and the outcome.

The level of abstraction can range from case-specific causal mechanisms, which are highly specified mechanisms that describe how a causal process works in a particular case, over

Table 14.1 *Comparing narratives to develop ideas about mechanisms*

Case A	Cause	Event 1a	Event 2a	Event 3a	Event 4a	Event 5a	→ Outcome
Case B	Cause	Event 1b	Event 2b	Event 3b	Event 4b	Event 5b	→ Outcome
Commonalities?	Cause	→ part 1 ($e_1 * a_1$)		→ part 2 ($e_2 * a_2$)		→ part 3 ($e_3 * a_3$) → Outcome	

contingent mechanisms that can be present in at least two cases, and mid-range mechanisms, which are processes described in quite abstract terms but still identify interlocking parts of the process in terms of entities engaging in activities. Entities and activities in a case-specific mechanism will typically be described using proper nouns (e.g. particular people or institutions), and the activities described will be the specific activities that they engage in. In contrast, in a mid-range mechanism the parts are quite abstract, focusing only on the most causally critical elements that are shared across a range of similar cases. Logically, when we lift the level of abstraction by dropping specifics, this expands the potential scope of generalizations about the mechanism, whereas lowering the level of abstraction involves further specification of parts of the process, thereby also narrowing the scope of potential generalization. But in all instances, to qualify as a mechanism-based explanation in a systems-understanding, we still have to be able to answer the “how does it work” question, which requires that the causal arrow in between causes and outcome has to be elucidated in enough detail that the critical parts of the process are made clear in terms of entities engaging in activities that link one part to the next (Craver and Darden 2013: 31; Hedström and Ylikoski 2010: 53).

If one is in the dark about the process, then a theory-building case study should be employed, which involves a back-and-forth between empirics and theory after a typical case is chosen (see below). One can start with an event chronology, but the core of theory-building is using existing theorization as a source of inspiration for what to look for when engaging in an abductive process aimed at detecting a plausible mechanism through the empirical traces it leaves.⁴ A particularly useful way of formulating mid-range theories of policy processes is to develop case-specific narratives of two cases that appear to be relatively similar in terms of causes, outcome, and context. Once narratives in terms of events are developed for both, one can engage in a process-focused comparison to see whether there are any commonalities that might reflect parts of a more abstract mechanism. For example, do we see particular types of social actors engaging in roughly analogous activities at similar times in the two cases? If so, this commonality might reflect an underlying part of a mechanism that might be present in the two cases. This comparative process can be continued until the mechanism is developed – either as a minimalist “sketch” or as a full-fledged mechanistic explanation that can then be tested more systematically. This is depicted in Table 14.1.

The next step involves a mapping of a population using exploratory comparative methods. This can be accomplished using a simple comparison of cases, or more systematic methods like Qualitative Comparative Analysis (QCA). Irrespective of whether the method is systematized (QCA) or more ad hoc, cases are compared on values of the cause(s), outcome, and potential contextual conditions that might matter for which process is operative. The result of the comparison can look like Table 14.2. Here we can see that there are three cases where the cause is present (1, 2, and 4), but only two cases where both cause and outcome are present (1, 2). Cases 1 and 2 appear similar, although they still might differ on unknown contextual condition, which means that there might be different processes operative linking cause and outcome even in these two cases.

Table 14.2 A simple mapping of a population

Case	Cause	Contextual condition 1	Contextual condition 2	Contextual condition 3	Outcome
1	+	+	+	-	+
2	+	+	+	-	+
3	-	-	+	-	+
4	+	-	-	-	-

Note: + means that condition is present, - means that condition is absent.

Table 14.3 Four types of cases

Outcome present	II – Deviant case (coverage) (used to find new causes)	I – Typical cases (policy process can in theory be present)
Outcome not present	III – Irrelevant cases Cause(s) and/or contextual conditions not present	IV – Deviant case (consistency) (used to detect omitted conditions that led to process breakdown) Cause(s) and known contextual conditions present

Mapping a population enables us to group cases into sets that at the cross-case level appear to be relatively causally homogeneous (same values of causes, outcomes, and most important contextual conditions). Four types of cases result from the comparative mapping: positive (typical) cases, deviant consistency (where the relationship should have occurred but it did not), deviant coverage (other causes produce the outcome), and irrelevant cases (negative on cause(s) and outcome). Negative cases are irrelevant for studying policy processes because there is no process to trace. The four types of cases and what uses they have are depicted in Table 14.3.

Once the population is mapped, one can choose a typical (positive) case to engage in either testing of the theorized mechanism, or theory-building (see above).

If one is testing a mechanism, it is only after a case has been selected that one can develop propositions about observable manifestations – aka empirical fingerprints – that might be left by the activities of entities in the case. The reason for this is that the fingerprints left by activities vary depending on the case; the same empirical trace can mean something different in different evidential settings.

The link between the theoretical level (causal mechanism) and actual empirical observations has two steps: one theoretical, the other more empirical (see Figure 14.1) (Beach and Pedersen 2019: 155–222). Building on Bayesian logic, the first step involves theorizing about what potential observable manifestations an activity of an entity might have in a given case – what can be thought of as the empirical fingerprints of an activity.

Bayesian logic suggests that we have to answer two questions in relation to operationalizing empirical fingerprints: do we have to find it (*theoretical certainty*), and if we find it, are there alternative plausible explanations for finding the fingerprint other than that the activity happened (*theoretical uniqueness*)? If a fingerprint is theoretically certain and we do not find it, this enables us to disconfirm that the activity took place to a degree contingent upon how certain it was. If a fingerprint is theoretically unique and we find it, this enables confirmation to a degree contingent upon how unique it is. Theoretically unique evidence of an activity can be thought of as analogous to direct evidence in a criminal trial. For example, if we are

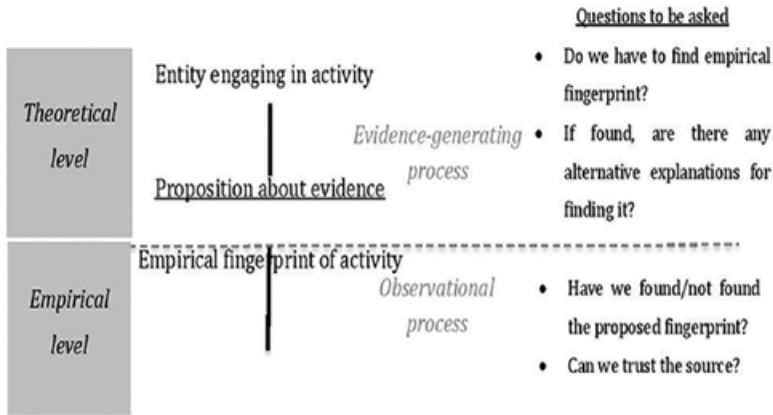
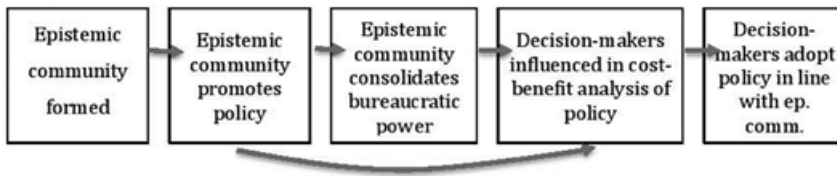


Figure 14.1 A two-stage evidence-evaluation framework for turning empirical material into evidence of mechanisms

tracing a policy diffusion process, a part of the process might be that actors in country A seek inspiration for how to design a policy by looking at an adopted policy in country B. A potential fingerprint of this activity could be that the subsequent policy in country A has a number of significant similarities with that in country B that can act as a policy signature. However, a more theoretically unique fingerprint would be if this policy signature includes elements that were developed to deal with the particular context in country B, but that do not really match the context in country A. If we find this type of non-adapted policy signature in country A, this would be quite direct evidence that policy makers were inspired by country B's policy.

However, *predictions* about empirical fingerprints are not yet evidence upon which causal inferences can be made. After operationalization, we have to explore the empirical record to assess whether the predicted empirical fingerprints were actually present or not. The questions we have to evaluate relate to *empirical certainty* or *empirical uniqueness*, depending on whether we find observations that we think are the predicted empirical fingerprints or not (Sober 2009: 68). If we find an observation, we have to ask ourselves whether the observation is really what we think it is and whether we can trust it. In effect, we are asking whether there are any credible empirical reasons for why the observation is not the proposed fingerprint (empirical uniqueness). If we do not find an observation, we have to ask ourselves whether we had access to the full empirical record or not (empirical certainty). For example, if we are using participant interviews, our sources might have had incentives to hide events from us. If we do not have full access, we cannot conclude that absence of evidence is evidence of absence.

The result of this two-stage process is mechanistic evidence that enables us to make causal inferences about the presence/absence of causal mechanisms. Assessing the full evidential picture for the activities associated with each part of the policy process involves evaluating the probative weight of found/not found confirming and disconfirming pieces of evidence. Often the initial mechanism sketch will be wrong, which should then lead to a theoretical revision of the mechanism. This means that within-case analysis in case-based research is a process of revising a process-theory, using empirical evidence to update one's beliefs about the process that was operative in a case.



Source: Löblová (2018: 165).

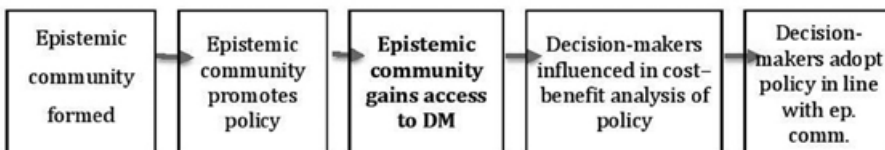
Figure 14.2 Löblová's initial mechanism

As an example, Löblová (2018) develops an initial mechanistic sketch that theorizes a process linking the formation of an epistemic community (cause) with decision-makers adopting a policy in line with what the epistemic community wants (outcome) (depicted in Figure 14.2).

After engaging in a case study, she finds that the process did not function exactly as hypothesized. She found in the case study that access was a critical step, and that merely promoting a policy is not enough to influence decision-makers. This resulted in a revised process-theory, depicted in Figure 14.3.

Just because one finds a policy process operative in one case, this does not mean that one can generalize that the same process should be operative in all other typical (positive) cases in the population. As discussed above, mechanisms at this level of abstraction are sensitive to contextual differences. Therefore, after tracing of a policy process in one case, it is imperative to explore whether similar processes are operative in cases that our comparison suggested are relatively similar. In Table 14.2, after finding enough mechanistic evidence of the process in case 1 that we are reasonably confident it operated as we hypothesized, one could then choose to trace the process in case 2, in which there are no known different contextual conditions.⁵ Tracing of a process in additional cases can be made easier if it is possible to focus on only one critical part of the process, or there is a highly probative signature trace of the particular process that can be searched for in other cases. If this is possible, then instead of tracing the full process, one can increase one's confidence that it is a similar process operative in other cases by only investigating the critical part/critical signature.

If one finds a similar process operative in case 2, we can then cautiously infer that the process should work in cases similar to 1 and 2. However, we cannot generalize to case 4, or other potential cases that differ on known contextual conditions. Therefore, we should explore even more dissimilar cases, engaging in a snowballing-outwards strategy of exploring the bounds of the operation of a particular policy process (Beach, Pedersen, and Siewert 2019).



Source: Löblová (2018: 176).

Figure 14.3 Löblová's revised mechanism

In parallel, it can be useful to trace mechanisms in deviant consistency cases, where the cause(s) and known contextual conditions are present, but the outcome does not occur in a situation where it should have. Existing theories of mechanisms that were either built or tested on typical cases provide the foundation for the tracing in deviant cases until mechanism breakdown. Tracing the mechanism until it breaks down helps shed light on omitted causal and/or contextual conditions. Finding out when and why a mechanism breaks down gives us clues about omitted contextual or causal conditions, although the process tracing component here is auxiliary, and the main analytical method used to detect omitted conditions is a systematic paired comparison of the deviant case with a typical case.

Taken together, a case-based approach to studying public policies aims to understand how policy processes work within cases. The analytical heavy-lifting is done by within-case tracing of policy processes in individual cases. Comparisons have the primary role of mapping populations of cases that enable the selection of appropriate cases for within-case analysis.

5. CONCLUSIONS

Case-based approaches to studying public policy have the goal of learning about how policy processes actually operate within real-world cases. The upside is that we learn a lot about how a policy intervention works in a given context, and the conditions required for it to work in a particular way. The downside is that our knowledge claims are relatively context-specific.

Many variance-based scholars are skeptical about research that makes this type of relatively particularistic, bounded inferences. Gerring writes that “social science gives preference to broad inferences over narrow inferences. First, the scope of an inference usually correlates directly with its theoretical significance. Second, broad empirical propositions usually have greater policy relevance, particularly if they extend to the future. They help us to design effective institutions. Finally, the broader the inference, the greater its falsifiability” (2017: 234). Scholars within the case-based approach counter that complexity and contextual sensitivity are key features of twenty-first-century science, seen in developments in fields like systems biology or systems/personalized medicine (Ahn et al. 2006; Bechtel and Richardson 2010; Cartwright 2007, 2012; Levi-Montalcini and Calissano 2006). Instead of research that aims to evaluate the effect of individual treatments in isolation across large heterogeneous populations, systems biology and personalized medicine seek to investigate how treatments work within subgroups of complex, real-world systems – in other words, small bounded populations of relatively similar cases. Appreciating complexity means that our claims become more contextually specific (Bechtel and Richardson 2010). Instead of engaging in a simple experiment that isolates the effect of a treatment in a controlled environment, researchers are increasingly interested in exploring how things work in particular contexts (Cartwright 2011, 2012). Applied to policy studies, this would mean that we learn about how a policy intervention works in a particular context by tracing the policy process triggered by it. However, we would not assume that the policy intervention would work in the same way (if at all) in other policy contexts. Instead of one-size-fits-all claims, case-based policy analysis tries to understand how things work in particular contexts.

Appreciating complexity does not mean that we cannot engage in cumulative research. Ideally, after intensive collaborative research over a longer time period, the result would be an evidence-based catalog of different mechanisms that are triggered by a given cause (or set of

causes) in different contexts. Naturally, this type of research demands more resources, but this is not an excuse to engage in sloppy generalizations about mechanisms.

Finally, Gerring's claim about policy relevance does not match recent developments in the field of policy evaluation, where there is increasing interest in using case-based methods to trace mechanisms as an analytical tool to study how interventions work in particular contexts *instead* of working with broad propositions that tell us little about how things work in the real world (Bamanyaki and Holvoet 2016; Befani and Stedman-Bryce 2016; Cartwright 2011; Clarke et al. 2014; Schmitt and Beach 2015; Wauters and Beach 2018).

Case-based approaches to policy analysis are still under development. Significant methodological progress has been made in the past decade, but much work still remains. In particular, while many of the ideas inspired by the mechanistic turn in the natural sciences have been incorporated into methodological guidelines, the practical utility of the tools for parts of the research process such as operationalizing mechanisms and generalization after case studies has to be explored more.

NOTES

1. Within medicine, the top of the hierarchy is meta-studies that evaluate the findings of multiple RCTs on the same research question in different settings.
2. If the outcome is unknown, a preliminary tracing should be undertaken to find an outcome of interest.
3. For more on how to conceptualize concepts in case-based research, see Beach and Pedersen (2016: chapters 4 and 5).
4. For more, see Beach and Pedersen (2019: chapter 9); Swedberg (2014).
5. There can of course always be unknown omitted conditions that might also affect how a process works.

REFERENCES

- Adcock, R. (2007). Who's afraid of determinism? The ambivalence of macro-historical inquiry. *Journal of the Philosophy of History* 1: 346–64.
- Ahn, A. C., Tewari, M., Poon, C.-S., and Phillips, R. S. (2006). The limits of reductionism in medicine: could systems biology offer an alternative? *PLoS Medicine* 3 (6): e208. <https://doi.org/10.1371/journal.pmed.0030208>.
- Angrist, J. A. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Bamanyaki, P. A. and Holvoet, N. (2016). Integrating theory-based evaluation and process tracing in the evaluation of civil society gender budget initiatives. *Evaluation* 22 (1): 72–90.
- Beach, D. and Pedersen, R. B. (2016). *Causal Case Studies: Foundations and Guidelines for Comparing, Matching, and Tracing*. Ann Arbor, MI: University of Michigan Press.
- Beach, D. and Pedersen, R. B. (2019). *Process-Tracing Methods: Foundations and Guidelines* (2nd edition). Ann Arbor, MI: University of Michigan Press.
- Beach, D., Pedersen, R. B., and Siewert, M. (2019). Case selection and nesting of process-tracing studies. In D. Beach and R. B. Pedersen, *Process-Tracing Methods: Foundations and Guidelines* (2nd edition). Ann Arbor, MI: University of Michigan Press, 89–154.
- Bechtel, W. and Richardson, R. C. (2010). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Cambridge, MA: MIT Press.
- Befani, B. and Stedman-Bryce, G. (2016). Process tracing and Bayesian updating for impact evaluation. *Evaluation* 23 (1): 42–60.

- Bhaskar, R. (1979). *A Realist Theory of Science*. Brighton: Harvester.
- Bunge, M. (1997). Mechanism and explanation. *Philosophy of the Social Sciences* 27 (4): 410–465.
- Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
- Cartwright, N. (2007). *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge: Cambridge University Press.
- Cartwright, N. (2011). Predicting “it will work for us”: (way) beyond statistics. In P. M. Illari, F. Russo, and J. Williamson (Eds.), *Causality in the Sciences*. Oxford: Oxford University Press, 750–68.
- Cartwright, N. (2012). Will this policy work for you? Predicting effectiveness better: how philosophy helps. *Philosophy of Science* 79 (5): 973–89.
- Clarke, B., Gillies, D., Illari, P., Russo, F., and Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi* 33 (2): 339–60.
- Craver, C. F. and Darden, L. (2013). *In Search of Mechanisms*. Cambridge: Cambridge University Press.
- Dowe, P. (2011). The causal-process-model theory of mechanisms. In P. M. Illari, F. Russo, and J. Williamson (Eds.), *Causality in the Sciences*. Oxford: Oxford University Press, 865–79.
- Falletti, T. G. and Lynch, F. (2009). Context and causal mechanisms in political analysis. *Comparative Political Studies* 42: 1143–66.
- Gerring, J. (2010). Causal mechanisms: yes but ... *Comparative Political Studies* 43 (11): 1499–526.
- Gerring, J. (2011). *Social Science Methodology: A Unified Framework*. Cambridge: Cambridge University Press.
- Gerring, J. (2017). *Case Study Research* (2nd edition). Cambridge: Cambridge University Press.
- Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis* 44 (1): 49–71.
- Goertz, G. and Levy, J. S. (Eds.) (2007). *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*. London: Routledge.
- Goertz, G. and Mahoney, J. (2009). Scope in case-study research. In D. Byrne and C. C. Ragin (Eds.), *The SAGE Handbook of Case-Based Methods*. Thousand Oaks, CA: Sage, 307–17.
- Goertz, G. and Mahoney, J. (2012). *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.
- Groff, R. (2011). Getting past Hume in the philosophy of social science. In P. M. Illari, F. Russo, and J. Williamson (Eds.), *Causality in the Sciences*. Oxford: Oxford University Press, 296–316.
- Hedström, P. and Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology* 36: 49–67.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* 81 (396): 945–60.
- Illari, P. M. (2011). Mechanistic evidence: disambiguating the Russo-Williamson thesis. *International Studies in the Philosophy of Science* 25 (2): 139–57.
- Illari, P. M. and Williamson, J. (2011). Mechanisms are real and local. In P. M. Illari, F. Russo, and J. Williamson (Eds.), *Causality in the Sciences*. Oxford: Oxford University Press, 818–44.
- King, G., Keohane, R. O., and Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Leamer, E. E. (2010). Tantalus on the road to asymptopia. *Journal of Economic Perspectives* 24 (2): 31–46.
- Lebow, N. (2000). Contingency, catalysts and international system change. *Political Science Quarterly* 115: 591–616.
- Levi-Montalcini, R. and Calissano, P. (2006). The scientific challenge of the 21st century: from a reductionist to a holistic approach via systems biology. *BMC Neuroscience* 7 (Suppl. 1): S1. doi:10.1186/1471-2202-7-S1-S1.
- Levy, J. (2015). Counterfactuals, causal inference, and historical analysis. *Security Studies* 24 (3): 378–402.
- Löblová, O. (2018). When epistemic communities fail: exploring the mechanism of policy influence. *Policy Studies Journal* 46 (1): 160–189.
- Machamer, P. (2004). Activities and causation: the metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science* 18 (1): 27–39.
- Machamer, P., Darden, L., and Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science* 67 (1): 1–25.

- Mahoney, J. (2008). Toward a unified theory of causality. *Comparative Political Studies* 41 (4–5): 412–36.
- Morgan, S. L. and Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press.
- Ragin, C. C. (1987). *The Comparative Method Moving Beyond Qualitative and Quantitative Strategies*. Berkeley, CA: University of California Press.
- Ragin, C. C. (2000). *Fuzzy-Set Social Science*. Chicago, IL: University of Chicago Press.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review* 15 (3): 351–7.
- Rubin, D. (2005). Causal inference using potential outcomes: design, modeling, decisions. *Journal of the American Statistical Association* 100 (469): 322–31.
- Runhardt, R. W. (2015). Evidence for causal mechanisms in social science: recommendations from Woodward's manipulability theory of causation. *Philosophy of Science* 82 (5): 1296–307.
- Russo, F. and Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science* 21 (2): 157–70.
- Russo, F. and Williamson, J. (2011). Generic versus single-case causality: the case of autopsy. *European Journal of the Philosophy of Science* 1 (1): 47–69.
- Schmitt, J. and Beach, D. (2015). The contribution of process tracing to theory-based evaluations of complex aid instruments. *Evaluation* 21 (4): 429–47.
- Schneider, C. Q. and Rohlfing, I. (2013). Combining QCA and process tracing in set-theoretical multi-method research. *Sociological Methods and Research* 42 (4): 559–97.
- Schneider, C. Q. and Rohlfing, I. (2016). Case studies nested in fuzzy-set QCA on sufficiency: formalizing case selection and causal inference. *Sociological Methods and Research* 45 (3): 526–68.
- Seawright, J. (2016). *Multi-Method Social Science*. Cambridge: Cambridge University Press.
- Sober, E. (2009). Absence of evidence and evidence of absence: evidential transitivity in connection with fossils, fishing, fine-tuning, and firing squads. *Philosophical Studies* 143 (1): 63–90.
- Swedberg, R. (2014). *The Art of Social Theory*. Princeton, NJ: Princeton University Press.
- Tetlock, P. E. and Belkin, A. (Eds.) (1996). *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*. Princeton, NJ: Princeton University Press.
- Waskan, J. (2011). Mechanistic explanation at the limit. *Synthese* 183 (3): 389–408.
- Wauters, B. and Beach, D. (2018). Process tracing and congruence analysis to support theory-based impact evaluation. *Evaluation* 24 (3): 284–305.
- White, H. (2009). Theory-based impact evaluation: principles and practice. International Initiative for Impact Evaluation Working Paper 3, New Delhi, India.
- Williams, M. and Dyer, W. (2009). Single case probabilities. In D. Byrne and C. C. Ragin (Eds.), *The SAGE Handbook of Case-Based Methods*. Thousand Oaks, CA: Sage, 84–100.
- Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

15. Qualitative Comparative Analysis for comparative policy analysis

Eva Thomann

1. INTRODUCTION

This chapter¹ discusses different approaches to Qualitative Comparative Analysis (QCA) and their usefulness in addressing major themes of comparative public policy analysis in different research contexts. There seems to be a preferential connection between QCA and public policy analysis both in terms of research design and in terms of the actual needs and goals of policy-oriented research (Rihoux, Rezsöhazy, and Bol 2011; Rihoux et al. 2013). At the same time, recent years have witnessed a diversification of research approaches to QCA. In many ways, QCA, which was always intended to be a multi-method approach, is situated in between quantitative and qualitative methodologies (Ragin 2014 [1987]). Accordingly both small-N and large-N, exploratory and theory-led variants of QCA have developed, which prioritize either the parsimony of the results or their substantive interpretability (Thomann and Maggetti 2017). While this diversification has broadened our understanding of the nature, purpose, and usefulness of QCA, it can also create misunderstandings and provoke criticisms (Tanner 2014), thus providing a welcome opportunity for self-reflection and improvement.

This chapter asks the following question: What do different approaches to QCA offer to comparative public policy analysis? To answer this question, this chapter first briefly introduces the core features of the QCA method that influence how well it is suited to public policy analysis. Second, the chapter discusses how these core features relate to seminal perspectives, research questions, and analytic goals in comparative policy analysis. Third, the chapter introduces a recently developed typology of QCA approaches (Thomann and Maggetti 2017) which differ in three ways, and critically discusses the strengths, limitations, and implications for public policy research. Recently published QCA applications illustrate how different QCA approaches can inform public policy scholars when designing their research. A key message is that the particular assumptions underlying QCA set out the many possibilities for its use in small-N and large-N, theory-developing, and theory-evaluating research designs, focusing on different aspects of causation. The chapter identifies situations in which different approaches to QCA may be helpful for public policy analysts.

2. QCA AS A TECHNIQUE

When situating QCA among other methods of comparative public policy, it is useful to distinguish QCA as a technique from QCA as an approach. As Thomann and Maggetti (2017) discuss, the QCA technique refers to formalized data analysis based on data-set observations involving truth table analysis and logical minimization (Rihoux and Ragin 2009). The QCA technique is unique in that it combines an analysis of complexity – preserving cases as config-

urations of attributes – with a systematic cross-case comparison in order to detect regularities (Berg-Schlosser et al. 2009; Engeli, Rihoux, and Rothmayr Allison 2014).

2.1 Assumptions of QCA

Thomann and Maggetti (2017) describe how explanatory uses of QCA can address research questions concerning the *causes of a given effect* in terms of set relations, that is (quasi-)necessary and/or (quasi-)sufficient conditions, and assuming complexity (Mahoney and Goertz 2006). A condition X is necessary (\leftarrow) for an outcome Y if whenever Y is given, X is also given (that is, Y implies X; and, Y is a subset of X). X is sufficient (\rightarrow) for Y if whenever X occurs, Y also occurs (that is, X implies Y; and, X is a subset of Y). QCA specifically models three aspects of *complexity*: asymmetry, equifinality, and conjunctural patterns (Berg-Schlosser et al. 2009; see also Mahoney, Kimball, and Koivu 2009). Asymmetry means that the conditions leading to the occurrence of an outcome can differ from those leading to its non-occurrence. Equifinality relates to the fact that many roads can lead to Rome: the same phenomenon can have different, mutually non-exclusive explanations. When focusing on conjunctural patterns, the researcher does not assume isolated effects of single variables. Instead the effect of a single condition might unfold only in combination with other conditions (Schneider and Wagemann 2012: 78; see also Baumgartner and Thiem 2017b; Thiem, Baumgartner, and Bol 2016). For instance Damonte (2014) argues that mainstream assumptions about how environmental effectiveness results from interests and policy making have produced erratic results. Instead she uses QCA to explain effectiveness through the interplay between policy tools and ideas (conjunctural patterns).

2.2 Steps of a QCA Analysis

The typical QCA analysis proceeds as follows (described in Oana, Schneider and Thomann 2020; for comprehensive introductions, see, for example, Ragin 2008b; Rihoux and Ragin 2009; Schneider and Wagemann 2012). The selection, conceptualization and operationalization of cases, the outcome (explanandum) and conditions (explanans) follow the protocols of (qualitative) comparative research design and measurement validity (Adcock and Collier 2001; Radaelli and Wagemann 2019; Toshkov 2016). QCA is distinctive in that it requires an additional step in which social phenomena are translated into sets through a process called calibration (De Block and Vis 2019; Ragin 2008a). QCA can handle different types of sets in the same analysis, reflecting the nature of the underlying phenomena: dichotomous crisp sets (Ragin 2014 [1987]), differences in degree as expressed in fuzzy sets (Ragin 2000, 2008b), and polytomous multi-value sets (Cronqvist and Berg-Schlosser 2009; Haesebrouck 2015). All sets establish a difference in kind between partial or full membership in the set (for example, partially or fully successful implementation), and partial or full non-membership in the set (for example, partially or fully unsuccessful implementation) (see also Duşa 2018; Thiem 2014).

QCA integrates parameters of fit to assess how ‘perfect’ (consistency) and empirically relevant (coverage) a set relation is (Thomann 2019). *Consistency* is the extent to which the results are in line with the statements of necessity or sufficiency. *Coverage* tells us about the empirical importance of necessary and sufficient conditions. The analysis of necessity (Schneider 2018) often starts with identifying simple conditions that are a superset of (that is, they are necessary

for) the outcome. If no simple condition proves necessary, further simple conditions can be added disjunctively until necessity is obtained (Duşa 2018).

The analysis of sufficiency is based on the so-called *truth table* which depicts all logically possible configurations of conditions (Thomann 2019). We can insert the cases into the truth table rows and identify empirically unobserved configurations (so-called logical remainders). If the set membership of all or enough cases in a truth table row is smaller than or equal to its membership in the outcome, then the row is identified as a sufficient configuration for the outcome. The *logical minimization* process then serves to identify the shortest possible expression depicting the configurations that imply the outcome – the solution term. For example, let the Boolean multiplication sign * denote the logical AND, + the logical OR, and the tilde sign ~ the logical NOT. It is easy to see that $A*B*C + A*B*\sim C$ can be reduced to $A*B$ (for detailed descriptions, see Baumgartner and Thiem 2017a; Duşa 2018; Schneider and Wagemann 2012).

3. QCA IN COMPARATIVE PUBLIC POLICY ANALYSIS

The choice of a given method should follow the research question at hand (Schneider and Wagemann 2012; Toshkov 2016). There seems to be a clear relationship between prominent research questions that arise in public policy and the kind of patterns QCA enables researchers to identify (for example, Befani and Sager 2006; Engeli, Rihoux, and Rothmayr Allison 2014; Fischer and Maggetti 2017; Gerrits and Verweij 2016; Maggetti, Radaelli, and Gilardi 2012; Rihoux, Rezsöhazy, and Bol 2011; Varone, Rihoux, and Marx 2006). This theory-methods-fit has specific reasons and limitations (see Table 15.1) (Howlett, Ramesh, and Perl 1995; Jann and Wegrich 2007; Knill and Tosun 2012; Sabatier and Weible 2017).

3.1 Advantages of QCA for Public Policy Analysis

The first main advantage of QCA for public policy analysis is that the *notions of necessity and sufficiency can often accurately capture the analytic interests* of studying decision making, implementation, and evaluation outcomes (Rihoux, Rezsöhazy, and Bol 2011). These interests may include core requirements for achieving certain results (necessity). For example, public sector organizations might want to know what they need to do to ensure their employees are willing to implement a given policy (necessity). Or policy researchers focus on the situations in which particular outcomes – for example, the coupling of multiple streams – come about (sufficiency). In line with goal-oriented policy analysis, QCA techniques ‘allow policy analysts or evaluators to examine under which conditions a specific policy would be effective or not. ... QCA produces “deterministic” results that are applicable to groups or clusters of cases, in the form of: “this given combination of conditions leads to the outcome (say: a policy success) in such and such cases; by contrast, this other given combination of conditions does not lead to the outcome (say: a policy failure) in such and such cases”’ (Engeli, Rihoux, and Rothmayr Allison 2014: 89; see also Rihoux, Rezsöhazy, and Bol 2011).

The second main advantage of the QCA technique is that in *modelling complexity*, it ‘moves away, quite radically, from simplistic, probabilistic causal reasoning’ (Berg-Schlosser et al. 2009: 8–9; Ragin 2014 [1987]). Major theories of the policy process often include an assumption about complexity, especially in policy implementation and evaluation (Emmenegger, Kvist, and Skaaning 2013; Sabatier and Weible 2017; see also Rihoux, Rezsöhazy, and Bol

Table 15.1 Applying QCA to study policy processes

Process	Research questions, theories, and the strengths and limitations of the QCA technique
Problem definition and agenda-setting	<p style="text-align: center;"><i>Research questions</i></p> <p style="text-align: center;">Why do perceptions and definitions of policy problems change over time/vary from country to country? How do actors influence the definition of problems? Why are certain problems ignored while others are placed on the agenda?</p>
	<p style="text-align: center;"><i>Prominent theories and frameworks</i></p> <p style="text-align: center;">Multiple streams (Kingdon 1984), Punctuated equilibrium (Baumgartner and Jones 1993; True et al. 1999), Policy feedback (Pierson 1993)</p>
	<p style="text-align: center;"><i>Strengths</i></p> <p>Complexity: Models conjunctural coupling of multiple streams (Sager and Thomann 2017), contingent effects of issue attention and issue framing on the policy agenda (Dekker and Scholten 2017) Case sensitivity: Accounting for rare or non-events (ignored problems) through theory-guided case selection is possible (Goertz and Mahoney 2006)</p>
	<p style="text-align: center;"><i>Limitations</i></p> <p style="text-align: center;">QCA technique might not capture: – Changing agendas over time – Cognitive processes underlying the formation and change of perceptions and preferences</p>
Policy formulation and adoption	<p style="text-align: center;"><i>Research questions</i></p> <p style="text-align: center;">How do actors come to formulate policy solutions to policy problems? How do policy decisions come about? Why are some options preferred over alternative options? How can policy outputs, the way(s) in which they change over time, and variation outputs and their changes be explained?</p>
	<p style="text-align: center;"><i>Prominent theories and frameworks</i></p> <p style="text-align: center;">Advocacy Coalition Framework (Sabatier 1988), Institutional Analysis and Development Framework (Ostrom 2011), Policy diffusion (Braun and Gilardi 2006), Social construction of target groups (Schneider and Ingram 1993), Policy narratives (Jones and McBeth 2010)</p>
	<p style="text-align: center;"><i>Strengths</i></p> <p>Complexity: Models constellations of stakeholders and coalitions, the interplay between institutions and political interest representation, contextual scope conditions for causal relationships, combinations of target group characteristics, contextual contingency of narratives; non-decisions can have different explanations than decisions; several paths to policy decisions and change are possible Causes of effects: Identifies conditions that bring about policy decisions, outputs and change</p>
	<p style="text-align: center;"><i>Limitations</i></p> <p style="text-align: center;">QCA technique might not capture: – Changing explanations for policies over time – Processes underlying decision-making and narratives</p>

Process	Research questions, theories, and the strengths and limitations of the QCA technique
Policy implementation	<p style="text-align: center;"><i>Research questions</i></p> <p style="text-align: center;">Why do certain policies succeed at being carried out in practice and why do others fail? How do policies change when put into practice? Which factors account for variance in policy implementation? How do political, administrative, and societal actors interact in policy implementation?</p>
	<p style="text-align: center;"><i>Prominent theories and frameworks</i></p> <p style="text-align: center;">Conditions for effective implementation (Sabatier and Mazmanian 1980), Ambiguity-conflict model (Matland 1995), Forward and backward mapping (Elmore 1985), Integrated model of policy implementation (Winter 2003), Street-level bureaucracy (Lipsky 2010 [1980]), Multi-level governance (Hooghe and Marks 2003), Goodness of fit (Börzel and Risse 2003)</p>
	<p style="text-align: center;"><i>Strengths</i></p> <p>Set relations: Captures necessary and sufficient conditions (for example, capacity/willingness) for compliance/successful implementation</p> <p>Complexity: Captures the interplay between multiple design/institutional/organizational characteristics with domestic politics and constellations of actors, allows for more than one path to successful implementation contingent on idiosyncratic context, captures multi-level or fit-misfit configurations</p> <p>Causes of effects: The main goal of research is often to identify how successful implementation can be guaranteed; complex interplay of variables more insightful than isolated effects of causes</p>
	<p style="text-align: center;"><i>Limitations</i></p> <p style="text-align: center;">The QCA technique might not capture: – Processes underlying behaviour of implementing agents</p>
Policy evaluation and termination	<p style="text-align: center;"><i>Research questions</i></p> <p style="text-align: center;">How can policy effects (outcomes and impacts) be identified and improved? Under what conditions does the policy (not) achieve its goals? What are the effects of different policy measures? Which factors explain variation in policy effects?</p>
	<p style="text-align: center;"><i>Prominent theories and frameworks</i></p> <p style="text-align: center;">Context-mechanism-outcome (CMO) (Pawson and Tilley 1997), Utilization-focused evaluation (Patton 1997), Five criteria for policy evaluation (Knoepfel et al. 2011), Logic models (Funnell and Rogers 2011), Regulatory impact assessment (Radaelli and De Francesco 2010)</p>
	<p style="text-align: center;"><i>Strengths</i></p> <p>Set relations: Captures complex necessary and sufficient conditions (for example logic models) for goal achievement (outcomes and impacts)</p> <p>Complexity: Complexity is a core assumption of evaluation theories; Captures CMO configurations, input-output-outcome-impact configurations, allows for several paths to successful goal achievement, different explanations for success and failure possible; average effects often irrelevant for practitioners who seek to achieve certain results</p> <p>Causes of effects: The main goal of research is often to identify how goal achievement can be achieved, or how certain outcomes can be prevented</p> <p>Case sensitivity: Identifies pathways to success even if rare</p>
	<p style="text-align: center;"><i>Limitations</i></p> <p style="text-align: center;">Inherent limits in isolating causes of policy impacts QCA technique does not capture: – The extent of the impact of a certain measure – Processes underlying behavioural change of target groups</p>

Source: Own illustration drawing from Knill and Tosun (2012); Sabatier and Weible (2017).

2011: 15). In bringing about behavioural change, factors such as polity, policy, and politics never act in isolation, or as mutually independent variables (Sager and Andereggen 2012). Instead QCA can produce empirically well-grounded, context-sensitive evidence about policy instruments (Befani and Sager 2006; Pattyn, Molenveld, and Befani 2019; Rihoux, Rezsöhazy, and Bol 2011). Rather than following a ‘standard recipe’, policy effectiveness often depends upon a blend of unique ingredients, national/regional settings, sector-specific features, and cultural, political, and administrative traditions (Engeli, Rihoux, and Rothmayr Allison 2014: 89; Patton 1997; Thomann 2018). Identifying conjunctural patterns helps formulate useful policy recommendations, such as appropriate ‘policy mixes’ to achieve a given policy goal (Engeli, Rihoux, and Rothmayr Allison 2014: 88; for example, Thomann 2018); or they capture the intersectionality characterizing disadvantaged target groups (Ragin and Fiss 2017; Rihoux, Rezsöhazy, and Bol 2011: 56; Schneider and Ingram 1993). According to realistic evaluation approaches, a political programme can result in different outcomes depending on the context and assuming equifinality (Befani and Sager 2006; Falletti and Lynch 2009; Gerrits and Verweij 2018; Pawson and Tilley 1997).

The third main advantage of QCA is that it allows for the *systematic analysis of case study material within a quasi-experimental design* – arguably a typical research setting in policy analysis. The objects of interest for policy researchers and practitioners are often “‘naturally” limited in number: nation states or regions, different kinds of policies in different states, policy outputs and outcomes, policy programmes, policy styles, policy sectors, etc.’ (Rihoux, Rezsöhazy, and Bol 2011: 17). QCA offers procedures for systematic comparison of the case study material (for example, policy programmes) in a small- or medium-N design, while aiming to achieve a middle-range generalization. It allows for cross-national, cross-regional and cross-sectoral comparisons of macro-level (such as countries), meso-level (such as collective organizations), and micro-level (such as policy players) phenomena, nested within multi-level systems such as the European Union (EU), the Organisation for Economic Co-operation and Development (OECD), regions, countries and municipalities (Engeli, Rihoux, and Rothmayr Allison 2014: 89). In contrast with classic comparative case studies, this versatility of the QCA technique ‘opens up the possibility of achieving more parsimonious explanations for qualitative comparative research on a larger number of cases ... and so places QCA into a more cumulative knowledge research approach’ (Engeli, Rihoux, and Rothmayr Allison 2014: 86–7).

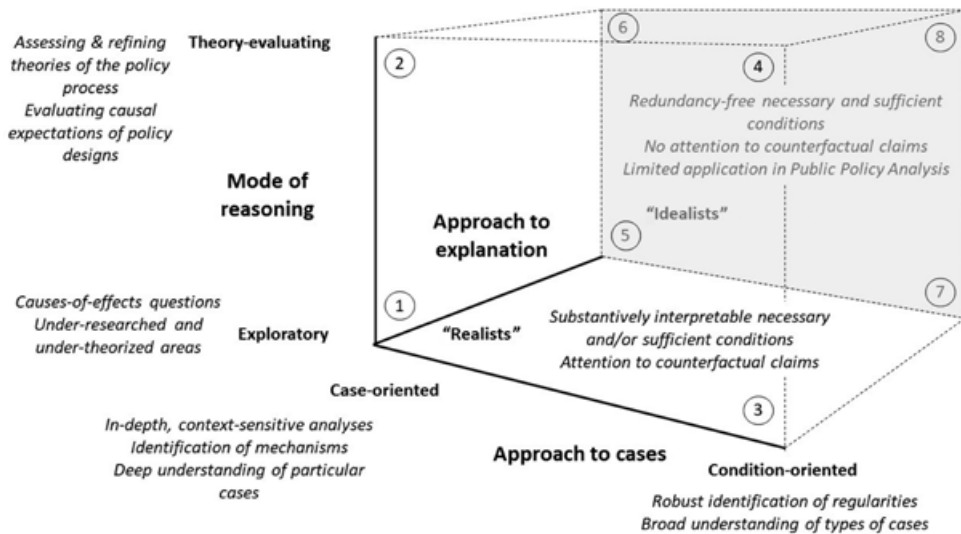
3.2 Limitations of QCA for Public Policy Analysis

One limitation of the cross-sectional QCA technique is that it does not easily deal with dynamic elements of the processes underlying policy outcomes and their evolution over time (Fischer and Maggetti 2017). Multi-method approaches to QCA help integrate dynamic elements. Recent advances include temporal QCA (Caren and Panofsky 2005), panel-data QCA (García-Castro and Ariño 2016; Hino 2009), and the systematic integration of process tracing through set-theoretic multi-method research and sequence analysis (Rohlfing and Schneider 2013, 2016; Williams and Gemperle 2017). Moreover, QCA is case-sensitive and not a suitable method for identifying average effect sizes (Tanner 2014) such as the precise effect size of a reduction of the education budget on disadvantaged target groups (Knoepfel et al. 2011; Radaelli and De Francesco 2010). Deriving general theories about policy phenomena or broad generalizations about populations is not usually possible due to the case-sensitivity and contin-

gency of QCA results. Instead QCA is best located in the more general area of ‘middle-range’ theorizing in social research (Befani 2013; Berg-Schlosser et al. 2009).

4. APPROACHES TO QCA IN COMPARATIVE PUBLIC POLICY

As Thomann (2019) highlights, QCA does not allow researchers to simply input data and let software find the solution (Berg-Schlosser et al. 2009). Rather QCA demands researchers be accountable and transparent about the choices they make regarding selecting and processing cases and variables, choosing tools, intervening during the analysis, and engaging in causal interpretations and generalization. Much of what makes a ‘good’ QCA relates to these processes. QCA as an approach includes fundamental questions of research design, that is, ‘the processes before and after the analysis of the data, such as the (re-)collection of data, (re-)definition of the case selection criteria, or (re-)specification of concepts’ (Schneider and Wagemann 2012: 11; see Peters 1998; Thomann and Maggetti 2017). QCA is inherently a multi-method approach (Ragin 2014 [1987]; Rihoux, Rezsöházy, and Bol 2011: 55). Thomann and Maggetti (2017) note that many current QCA uses do not seem to align with the original, case-oriented, inductive approach to QCA. There are now several approaches to QCA, which differ in three main ways: in the approach to cases, explanation, and modes of reasoning. The remainder of this chapter describes these approaches as introduced by Thomann and Maggetti (2017). It discusses their application to the comparative analysis of public policy using examples of studies published since 2014 (see Figure 15.1).



Source: Adapted from Thomann and Maggetti (2017) and Schneider (2018).

Figure 15.1 Approaches to QCA in comparative policy analysis

5. A CASE-ORIENTED OR CONDITION-ORIENTED APPROACH TO CASES

All QCA studies are configuration-oriented since cases are conceived of as a holistic configuration of attributes (Berg-Schlosser et al. 2009; Rihoux 2013). However, QCA approaches differ in the emphasis researchers put on cases as an object of inquiry (Greckhamer, Misangyi, and Fiss 2013; Thomann and Maggetti 2017).

5.1 The Case-Oriented Approach

As Thomann and Maggetti (2017) outline, the focus of the traditional case-oriented approach is close analysis of particular cases using deep contextual knowledge. In addition to cross-case inference, in-depth case knowledge helps establish measurement and internal validity. It emerges from an intensive qualitative engagement with the cases often based on purposively selected small- to medium-N samples (Berg-Schlosser and De Meur 2009). In-depth knowledge of cases, often acquired during primary data collection such as interviews and document analysis, helps researchers minimize measurement error *ex ante*, mitigate potential problems of limited diversity, and can clarify causally interpretable aspects of QCA results (Beach and Rohlfing 2018; Rohlfing and Schneider 2013; Schneider and Rohlfing 2013, 2016).

According to Thomann and Maggetti (2017), under such a case-oriented approach, cases are selected because obtaining in-depth knowledge about them is relevant for answering the research question (Mahoney and Goertz 2006). The theoretical criteria determining the relevance of the cases to the research question also constitute the scope conditions for the results; that is, the specific, explicitly defined empirical contexts in which observed relations or hypotheses apply (Byrne and Ragin 2009; Foschi 1997; Goertz and Mahoney 2006; Schneider and Rohlfing 2016). Case-oriented studies usually generate middle range rather than grand theories (Mahoney and Goertz 2006), which cannot be applied to other empirical contexts without further testing. This type of generalization is also called ‘limited’, ‘historical’, or ‘contingent’ (Blatter and Blume 2008; Rihoux and Ragin 2009).

5.2 The Condition-Oriented Approach

Thomann and Maggetti (2017) point out that condition-oriented applications focus on cases only in terms of a well-defined set of conditions. The results are interpreted mainly as patterns across cases and are not complemented with an in-depth, qualitative treatment of individual cases (Greckhamer, Misangyi, and Fiss 2013; Seawright and Collier 2010). This approach typically, though not necessarily, uses QCA on large samples, which are often implicitly or explicitly deemed representative. This facilitates resorting to complementary statistical techniques and parameters to evaluate QCA models (Cooper and Glaesser 2016; Fiss, Sharapov, and Cronqvist 2013; Greckhamer, Misangyi, and Fiss 2013). In the absence of qualitative case knowledge, and/or a priori guidance on the best model specification, a number of complementary strategies can assist to ensure measurement and internal validity (for example, Cooper and Glaesser 2016; Skaaning 2011). These strategies depend on the type of error that is expected to prevail (Maggetti and Levi-Faur 2013). According to Thomann and Maggetti (2017), a challenge for condition-oriented studies can be that inference is neither assisted by inferential statistics, nor by comprehensive case intimacy. A large N alone does not ensure

generalizability. A strong condition-orientation entails that researchers support their inferences with tools that typically work best on a relatively large number of cases (for example Braumoeller 2015; Eliason and Stryker 2009; Meuer, Rupietta, and Backes-Gellner 2016; Ragin 2000; Rohlfing 2018).

5.3 Discussion

Rather than the sheer number of observations, Thomann and Maggetti (2017) argue that what distinguishes these two approaches is the relative closeness to or distance from empirical cases. Whereas the case-oriented approach emphasizes the complementary use of within-case knowledge, the condition-oriented approach exclusively relies on cross-case inference, focused on relations between sets and based on knowledge of conceptual relationships rather than knowledge of particular cases. This orientation towards conditions is sometimes found in relatively small-N analyses. For example Jano (2016) analyses 34 cases in which pre-acceding southeast European countries did or did not comply with EU law. Despite the medium number of cases, the study relies little on case knowledge. Instead, it focuses on cross-case inferences which are tested extensively for robustness. Moreover a large number of cases does not preclude an interest in particular cases. Namely, set-theoretic multi-method research provides case selection criteria to assist in making inferences derived from cross-case comparisons with targeted within-case studies (Beach and Rohlfing 2018; Mikkelsen 2017; Rohlfing and Schneider 2013; Schneider and Rohlfing 2013, 2016). For instance Thomann (2018) analyses a relatively large set of 95 cases of customized implementation of EU food safety policies. The interpretation of results relies on in-depth studies of typical and deviant cases.

5.4 Usefulness for Public Policy Analysis

The *case-oriented approach* is useful for public policy scholars who want to perform in-depth analyses that will enable them to uncover causal mechanisms, and for scholars who work on subjects and questions where they expect the particular context to be highly relevant (Braun and Gilardi 2006; Falletti and Lynch 2009). For instance, Fischer (2015) uses QCA to study how institutions grant opportunities and impose constraints on actors when structures with a dominant coalition or with competing coalitions emerge. Fischer does this through an in-depth study and discussion of 11 important decision-making processes in Switzerland between 2001 and 2006.

Sager and Andereggen (2012; also Befani and Sager 2006) highlight that this case-oriented approach resonates particularly well with the assumptions of realist evaluation approaches which focus on the singularity and wholeness of each case (Pawson and Tilley 1997). This holistic, generative perspective is directly reflected in the way in which case-oriented QCA studies constantly use knowledge of individual cases in order to interpret the results (Sager and Andereggen 2012). Contingent generalization parallels the realist synthesis approach to generalization according to which the policy community ‘is not offered a “best buy” (approach “x” or case “y” seems to be the most successful) but a tailored, “transferable theory” (this programme theory works in these respects, for these subjects, in these kinds of situations)’ (Sager and Andereggen 2012: 64). In this vein Pattyn and Brans (2014) study the conditions that promote or impede the application of evaluation quality assurance mechanisms in 18 Flemish (Belgian) public sector organizations. Similarly, van der Heijden (2015) identifies three types

of roles governments play in the outcomes of voluntary environmental programs in Australia, the Netherlands, and the United States, and illustrates these roles using interview excerpts.

Conversely, the *condition-oriented approach* is less likely to be found in public policy analysis. This may partly be due to the prevalence of small- and medium-N data situations. Public policy scholars would choose a condition-oriented approach if the main goal of their analysis were to identify regularities that appear robustly in a range of cases. For example Thomann, van Engen, and Tummers (2018) evaluate central assertions of bottom-up implementation theory about the role of discretion in the willingness of frontline workers to implement public policies. Two large samples of Dutch street-level bureaucrats in two different policy sectors provide broad and robust support that perceived discretion is necessary for the motivation of policy implementers. Moreover, a condition-oriented approach enables scholars to gain a broad understanding of different *types* of policies, organizations, etc. For example, Boon and Verhoest (2014) explain reported overhead level by identifying three different types of agencies in terms of formal autonomy, result control, agency size, and task (N=44). Simultaneously, the absence of case knowledge poses significant challenges in performing the back-and-forth between ideas and evidence that QCA requires from analysts (see Wagemann, Buche, and Siewert 2016). Therefore case-oriented applications may remain the more attractive option for policy scholars.

6. A THEORY-GENERATING OR THEORY-EVALUATING MODE OF REASONING

Thomann and Maggetti (2017) also highlight that QCA studies differ in their modes of reasoning. The QCA technique can be fruitfully applied either to an exploratory, research design, or a confirmatory, research design (Eliason and Stryker 2009). Both modes of reasoning are valuable ways of contributing to knowledge and/or theory.

6.1 The Exploratory Approach

According to Thomann and Maggetti (2017), traditionally, QCA is often employed to ‘help the researcher generate some new insights, which may then be taken as a basis for a further theoretical development or for reexamination of existing theories’ (Berg-Schlosser et al. 2009: 16). Thus the bulk of QCA studies in public policy analysis adopt an exploratory approach that primarily aims to build or modify a hypothesis or abstract concept after the analysis (Rihoux, Rezsöházy, and Bol 2011). The studies start with data analysis from which specific conclusions or broader theoretical statements can be derived (Maggetti, Radaelli, and Gilardi 2012). QCA can be a powerful tool in generating set-theoretic hypotheses that account for causal complexity.

6.2 The Theory-evaluating Approach

In contrast to this traditional theory-generating approach, Thomann and Maggetti (2017) point out that QCA applications in public policy research increasingly explicitly formulate a priori expectations against which they compare their results. The primary aim of this theory-evaluating approach to QCA, which differs from traditional hypothesis testing, is to

evaluate existing, rather than to generate new knowledge. It starts out with an expectation which is then compared to and ultimately supported or refuted by empirical observations. This theory-evaluating approach to QCA is especially useful when a rich body of theoretical and substantial knowledge can be assessed and refined from a set-theoretic lens (Eliason and Stryker 2009). Expectations assessed with QCA must be formulated in line with set-relational patterns of (quasi-)necessity or (quasi-)sufficiency and/or aspects of complex causation (Fischer and Maggetti 2017; Schneider and Wagemann 2010, 2012; Thiem, Baumgartner, and Bol 2016).

Thomann and Maggetti (2017) argue that theory-evaluating QCA studies can inform us about the capacity, relevance, or relative strength of the theories used to explain and understand the case(s) under study, and that theory-evaluating QCA studies typically retain an iterative element (Blatter and Blume 2008). Formal set-theoretic theory evaluation, as developed by Ragin (2014 [1987]) and refined by Schneider and Wagemann (2012), is an especially interesting tool for this approach. It enables researchers to systematically evaluate set-theoretic propositions against the empirical results, based on the Boolean intersections of the solution terms, the theoretical propositions, and their negation. Other than traditional deductive hypothesis testing, this enables researchers to answer four questions. First, which parts of the hypothesis are supported by the findings? Second, in which directions should the hypothesis be expanded (exploratory)? Third, which parts of the hypothesis should be dismissed? Fourth, which cases are the most relevant for *ex post* within-case analysis? Schneider and Wagemann (2012) show how researchers can account for how many cases are members of the outcome and the non-outcome in the different intersecting areas (examples in Sager and Thomann 2017; Thomann 2015).

6.3 Discussion

Thomann and Maggetti (2017) note that while these two approaches adopt a different mode of reasoning, they correspond neither to the ideal notions of inductive designs, nor to the ideal notions of deductive research designs (Eliason and Stryker 2009). First, as an approach, QCA has an inherent iterative (or abductive) element that involves conceptual and theoretical considerations: researchers engage in a back-and-forth between prior knowledge and case knowledge. Theories, explanatory frameworks, concepts, and analytic decisions are refined based on preliminary empirical insights gained throughout the analysis; sampling and measurement decisions are re-specified using theoretical or conceptual insights (Berg-Schlosser et al. 2009; Schneider and Wagemann 2012). As deep theoretical knowledge should drive analytic decisions, ‘QCA is ill-equipped for analytic induction’ (Engeli, Rihoux, and Rothmayr Allison 2014: 88). Second, truth table analysis inherently entails a search for results, rather than simply testing the consistency and coverage of previously defined set-theoretic hypotheses (Thiem 2016a).

6.4 Usefulness for Public Policy Analysis

Exploratory QCA analyses are attractive for public policy scholars who operate in under-researched or under-theorized areas. As Gerring (2004: 349) notes, ‘path-breaking research is, by definition, exploratory’. Additionally, as Table 15.1 highlights, many prominent research questions in comparative public policy analysis are of the ‘causes of effects’ type

(Goertz and Mahoney 2006). Especially if the goal is to comprehensively understand why some hitherto unexplored outcome occurs, it is useful to be open to new, unexpected trajectories. For example, Cacciatore, Natalini, and Wagemann (2015) use a strongly exploratory approach to assess whether and to what extent different aspects of the EU2020 strategy have influenced the National Reform Programmes, resulting in patterns of ‘clustered Europeanization’. Hinterleitner, Sager, and Thomann (2016) explicitly prefer an exploratory approach to explain the International Monetary Fund’s evaluation of national austerity programmes, arguing that established theoretical approaches do insufficient justice to the context of the Eurozone.

More theory-led QCA applications are becoming increasingly common in public policy analysis (Rihoux, Rezsöhazy, and Bol 2011; Sager and Thomann 2017; Thomann 2015). This allows scholars to assess and refine the various theories of the policy process. For instance Pahl-Wostl and Knieper (2014) assess the ability of 27 water governance systems to deal with the climate change adaptation challenge. Their assessment is based on a strongly theory-led typology of the coordination and centralization of governance regimes. Shahidi (2015) uses QCA to explain the cross-national diversity of labour market policy responses to the Great Recession in 18 advanced welfare states. QCA enables them to combine the assertions made in a number of theoretical frameworks which attempt to explain cross-national patterns of welfare state recommodification in the aftermath of economic crises. Results suggest that ‘theories of welfare state change that attribute theoretical centrality to political and institutional factors do not provide a compelling explanation for patterns of labour market reform observed since the onset of the economic crisis’ (Shahidi 2015: 659).

As Table 15.1 illustrates, many public policy theories – for example, the multiple streams framework (MSF) (Kingdon 1984) or Sabatier and Mazmanian’s (1980) framework of necessary and sufficient conditions for successful policy implementation – implicitly or explicitly entail a set-theoretic, configurational logic or other aspects of causal complexity. Sager and Thomann (2017), for example, assess the applicability of MSF to explain differences in the labour market integration of asylum seekers in Swiss regions. Employing QCA enables them to assess the coupling of the problem stream, the policy stream, and the politics streams, and to integrate the role of institutional policy paths in the MSF.

A theory-driven observation is also a key feature of realistic evaluation approaches which emphasize the importance of building on theory in systematic review (Pawson and Tilley 1997). For practitioners ‘these techniques also allow one to test, both *ex post* and *ex ante*, alternative causal (policy intervention) models leading to a favorable/unfavorable policy output and favorable/unfavorable policy outcomes’ (Rihoux, Rezsöhazy, and Bol 2011: 16). The task of evaluation is to gather evidence to see if the process occurs as planned and, if it does not, to amend the theory to account for the divergent outcomes (Sager and Andereggen 2012). Gerrits and Verweij’s (2018) comprehensive guide for using QCA to evaluate complex infrastructure projects illustrates this.

7. EMPHASIZING SUBSTANTIVELY INTERPRETABLE OR REDUNDANCY-FREE MODELS

Finally, Thomann and Maggetti (2017) highlight the fact that there are different ideas about what makes a good and valid explanation using QCA. These questions become salient as limited empirical diversity – the fact that not all logically possible configurations are observed

in reality – focuses researchers' attention on the possibility of making counterfactual claims. Accordingly, two diverging protocols to analyse necessity and sufficiency have been put forward.

7.1 The QCA Realists

According to Thomann and Maggetti (2017), the traditional, more widespread approach emphasizes the *substantive interpretability* of QCA results from a practical research perspective in which social research 'is built upon a foundation of substantive and theoretical knowledge, not just methodological technique' (Ragin 2008b: 173). Proponents of this approach – Schneider (2018) calls them 'realists' – posit that a good explanation should be plausible and free of logical contradictions. Hence, the purpose of QCA is 'to find meaningful super- and/or subsets of the phenomenon to be explained' (Schneider 2016: 782). This approach views QCA results primarily as supersets or subsets of the outcome that differ in their complexity. Parsimonious and intermediate solution terms include configurations that were not empirically observed, but might occur in other settings. The problem with choosing counterfactual cases may be because there is a tendency to draw 'too many inferences on too little information' (Schneider and Wagemann 2016: 320; Wagemann, Buche, and Siewert 2016), or making inferences that are difficult to interpret.

When analysing sufficient conditions, QCA realists highlight that a parsimonious solution entails the assumption that all logical remainders that help eliminate redundancies are sufficient for the outcome – regardless of the 'goodness' of the counterfactual. In order to ensure accurate results, this approach requires that counterfactual claims be carefully justified (Emmenegger 2011). This can entail deriving either a complex (or conservative) solution that assumes that empirically unobserved configurations (logical remainders) are not sufficient for the outcome; or an intermediate solution based on carefully justified counterfactual arguments (Ragin 2008b; Schneider and Wagemann 2012, 2013, 2016). Directional expectations, based on theoretical and empirical knowledge, help distinguish plausible (easy) from implausible (difficult) counterfactuals (Standard Analysis, SA; Ragin 2008b). Untenable and other logically impossible arguments can be avoided through an appropriate treatment of remainders with Enhanced Standard Analysis (ESA) (Schneider and Wagemann 2016; for applications see Sager and Thomann 2017; Thomann 2015).

As Thomann and Maggetti (2017) explain, QCA realists interpret selected necessary conditions as crucial explanatory factors without which a given event could not have occurred (Goertz 2006; Goertz and Starr 2003; Schneider 2018; Schneider and Wagemann 2012). Supersets of the outcome can only be interpreted as conceptually meaningful necessary conditions if there are strong and plausible arguments that the conditions combined by the logical OR represent some higher-order construct; for example by operating as functional equivalents (Goertz and Mahoney 2005; Schneider and Wagemann 2010; 2012: 74). The empirical importance of necessary and sufficient conditions is assessed in a second analytic step: necessary conditions become empirically more important as they also approximate a sufficient condition, and sufficient conditions become more important as they approximate a necessary condition (Goertz 2006; Mahoney and Sweet Vanderpoel 2015; Schneider 2018).

7.2 The QCA Idealists

Thomann and Maggetti (2017) describe another, contrasting approach (the ‘idealists’ according to Schneider 2018) which highlights that ‘the crucial mechanism of QCA that turns necessary and sufficient conditions into causally interpretable necessary and sufficient conditions is the elimination of redundancies’ (Thiem and Baumgartner 2016a: 803). QCA idealists view QCA results primarily as causal claims. A ‘configurationally correct’ QCA solution only contains causally relevant factors (Baumgartner and Thiem 2017b). Such Boolean difference-makers are reliably revealed only by parsimonious solutions. While a host of supersets or subsets of an outcome exist, this approach derives causality only from conditions that are *both* minimally sufficient *and* contained in a minimally necessary condition for an outcome. That is, only a parsimonious solution that effectively eliminates all factors that are causally irrelevant (redundant) and has a very high coverage (indicating necessity) is causally interpretable (Baumgartner 2015; Baumgartner and Thiem 2017b).

According to this approach, intermediate and conservative solution formulas cannot be causally interpreted because they contain conditions that can be further eliminated (Baumgartner 2015). Hence different degrees of complexity are more than just a matter of specificity – they are about ‘false positives’. According to this approach, it is possible that an intermediate or conservative solution incorrectly attributes causal relevance to some factors, whereas the parsimonious solution does not (Baumgartner 2015). QCA idealists dissociate the concept of necessity without sufficiency from that of causality. Therefore, they consider it meaningless to propose criteria for the causal interpretation of necessary (or sufficient) conditions that are identified prior to minimization and are not redundancy-free (Thiem and Baumgartner 2016a).

7.3 Discussion

According to Thomann and Maggetti (2017), these two approaches present contrasting strategies to maintain internal validity when being faced with ‘noisy’ social science data (Schneider and Wagemann 2012). They also diverge on the existence of criteria that render (certain) necessary and sufficient conditions causally interpretable. Yet, both approaches refer to the INUS² theory of causation (Ragin 2000, 2008b; Schneider and Wagemann 2012; Thiem and Baumgartner 2016a; Thiem, Baumgartner, and Bol 2016). QCA realists and idealists also agree that ‘a set relation alone is not enough to postulate a cause’ (Schneider 2016: 782; Thiem, Baumgartner, and Bol 2016). The use of the QCA algorithm only describes super- and subset relations in the data, but does not provide a full explanation. That is, it does not provide a clarification of the mechanisms that explain why the (potentially) causal relationship holds.

At the time of writing, many QCA methodologists prefer the realist approach (Ragin 2008b; Rihoux and Ragin 2009; Schneider and Wagemann 2012; Schneider 2018). However, recently there has been some critique of the emphasis on substantive interpretability when engaging in causal interpretations. A recent study using simulated data sets posits that intermediate and conservative solutions sometimes produce ‘incorrect’ results, whereas parsimonious models are always correct (Baumgartner and Thiem 2017b). Written by ‘idealists’, this study defines ‘correctness’ as (amongst other things) the absence of causally irrelevant factors; only parsimonious solutions never contain causally irrelevant conditions. This methodological debate rests on divergent conceptions of what QCA results can or should tell us (see Schneider 2018).

Table 15.2 *Fictitious truth table*

Row	A	B	C	Outcome
1	1	1	1	1
2	0	0	0	0
3	0	0	1	0
4	0	1	0	0
5	0	1	1	?
6	1	1	0	?
7	1	0	1	?
8	1	0	0	?

Note: Bold: simplifying assumption for PS M1.

Although it is not possible to resolve this debate here, it is arguable that the approach emphasizing substantive interpretability usually resonates more with the research interests of public policy analysis. This argument relies on the premises that (a) typically there is limited diversity in public policy studies and (b) public policy scholars usually want to achieve some degree of, though modest, generalization beyond the analysed cases. A fictional, necessarily simplistic example may illustrate this. Let us assume that we use conditions A, B, and C to explain outcome Y, based on the truth table displayed in Table 15.2. The table contains one configuration that consistently results in outcome Y (row 1 with outcome value 1) and three configurations that do not consistently imply the outcome (rows 2–4 with outcome value 0). There are four logically possible combinations that we did not observe: rows five to eight marked with a question mark (“?”) in the outcome column (the so-called logical remainders). Let us also assume that we know the ‘true’ causally sufficient model (which in reality, with observational data, we never do):

M: $A*B \rightarrow Y$

Due to the presence of limited diversity, the QCA analysis might not be able to identify the true model M. Indeed, the truth table in Table 15.2 yields one conservative (CS) and two parsimonious solution models (PS):³

CS: $A*B*C \rightarrow Y$

PS M1: $A \rightarrow Y$

PS M2: $B*C \rightarrow Y$

The conservative solution CS was not able to detect that condition C is irrelevant for the production of outcome Y. This finding is an artefact of the available evidence, where all cases displaying A, B, and Y also display C. Crucially, QCA idealists consider the CS *incorrect* because the CS contains C even though C is, in truth, a causally irrelevant factor (Baumgartner and Thiem 2017b). According to QCA idealists this incorrect result arises because of the counterfactual assumption that all logical remainders are *not sufficient* for the outcome (outcome value 0). According to Baumgartner and Thiem (2017b), if confronted with imperfect data, QCA cannot be expected to produce ‘perfect’ or complete results. Instead *at least one of*

the identified models must not contain any causally irrelevant factors.⁴ PS M1 reads $A \rightarrow Y$. Under an approach emphasizing redundancy-free models, it is correct that condition A is *causally relevant* for outcome Y. Hence, PS M1 is correct whereas the CS is incorrect. In this approach, it does not matter that PS M1 is *incomplete* because it fails to detect the causal relevance of condition B; QCA idealists prioritize the principle of non-redundancy over the completeness of INUS configurations.

In contrast, QCA realists contend that in the true model M, A alone is not sufficient to reliably imply Y. The INUS condition A *has* to combine with the INUS condition B; this contingency is a core assumption of conjunctural causality. Hence, it seems short-sighted to claim that PS M2 is ‘correct’ under the assumption of INUS causation. For QCA realists, the empirical evidence in Table 15.2 alone does not suffice in concluding that B is irrelevant (as is C, for that matter). QCA realists contend that the PS M1 requires the (potentially illogical or untenable) counterfactual assumption that rows 6, 7, and 8 (had we observed them) would also have been sufficient for Y.⁵ This assumption produces an overly simplistic model. Finally, the laws of logic dictate that if $A*B$ consistently results in Y, then the combination $A*B*C$ (as well as any other subset of $A*B$) is also sufficient for the outcome Y. Therefore for QCA realists, the CS is entirely correct, though it may be overly specific. The conservative solution is simply the shortest possible description of the empirically observed sufficient configurations, which, absent measurement error, *cannot* contradict the ‘true’ model M.

7.4 Usefulness for Comparative Policy Analysis

Policy scholars and practitioners face a trade-off between ensuring the completeness of INUS configurations at the cost of allowing for redundancies, and ensuring causal relevance, at the cost of potentially incomplete INUS configurations.

Parsimonious models may omit indispensable parts of the ‘policy mix’ that consistently prevents or enables an outcome. This may, for example, potentially result in recommendations for policy makers to focus on measures which, on their own, are not effective (Zhang 2017). Generally the approach emphasizing substantive interpretability seeks to ensure that no causally relevant element is missing in the sufficient configurations. Naturally this aim requires QCA realists to avoid making counterfactual simplifying assumptions if there is neither empirical evidence nor a very plausible theoretical argument that a condition is irrelevant. For example Pattyn and Brans (2014: 370) outline that ‘given the lack of available theoretical and empirical evidence to make plausible assumptions, we predominantly rely, however, on the type of complex solution’. Intermediate solutions are designed to avoid both over- and under-simplification (Ragin 2008b). The limitation that some remaining elements could prove redundant should be made transparent. QCA realists are careful about engaging in causal interpretations; they often use case studies to shed light on underlying mechanisms and causal relevance (Schneider and Rohlfing 2013). For example Hooijer and Picot (2015) use the intermediate solution to analyse the institutional determinants that disadvantage immigrants in terms of poverty because it makes only relatively uncontroversial theoretical assumptions. They complement their fsQCA with three short case studies to examine the causal mechanisms.

Conversely, the approach emphasizing redundancy-free models can be suitable if researchers seek to identify only conditions that are causally relevant, without bothering about whether these configurations are, in the ‘true’ solution, sufficient for the outcome *on their own*.

Naturally, QCA idealists seek to avoid counterfactual claims that make a redundant condition appear causally relevant. This can be justified, for example, when deciding whether or not to invest public money in policy measures. For example, Nieto Morales, Wittek, and Heyse (2015) begin with the parsimonious solution, asking to what degree the availability of financial resources, in combination with other organizational characteristics, is a necessary precondition for compliance with reforms in governmental agencies. However, they then argue that the findings have no empirical foothold in their data and that they are difficult to defend from a theoretical point of view. It is worth noting that in the absence of the complete ‘policy mix’, the measures identified as causally relevant in the parsimonious model might be ineffective *in any context other than the one analysed in the given QCA study*. The purpose of the approach may be defeated if potentially relevant conditions are eliminated from the solution formula since causal relevance is identified, but not necessarily (causal) sufficiency beyond the sample at hand (Zhang 2017). This can be a limitation when there is a practical research expectation to interpret QCA solutions as potentially transferable sufficient configurations for a (policy) outcome. Transparency about this limitation (for example by formulating strict scope conditions) is advisable.

8. CONCLUSIONS

For policy researchers interested in the analysis of necessary and sufficient conditions and in modelling complexity, QCA offers a remarkably flexible tool for pursuing diverse analytical interests. Conversely, QCA struggles to model dynamic processes over time as well as average effect sizes. Thomann and Maggetti’s (2017) typology of QCA approaches allows public policy scholars to identify when to use QCA for the purposes of generating or evaluating theories; using in-depth within-case knowledge, or focusing on ‘types’ of cases; focusing either on the parsimony or on the completeness of sufficient configurations.

In real-life research settings where the ‘true’ result is always unknown, QCA has limitations, much like any other method for empirical cross-case comparison:

Of course, QCA techniques do not guarantee the final grasp of the ‘true’ causal grounds of a given phenomenon because the issue of causality is a much more complex matter ... Yet, if several competing theories try to explain the same result, QCA techniques will quickly disqualify the theories that are unable to discriminate correctly between cases with and without the outcome under study ... among the remaining theories, those that best satisfy the ‘parsimony principle’ (Occam’s ‘razor’) will emerge. ... as Einstein put in his famous dictum: One should express things ‘as simply as possible, but no simpler’. (Berg-Schlosser et al. 2009: 10)

As an inherent and unavoidable aspect of reality, limited diversity can cause QCA results to err in different directions (see also Braumoeller 2015; Rohlfing 2018). However, it is important to note that all QCA solution types contain an accurate description of the sufficiency patterns in the given data set. Arguably, it is this observation that informs the advice of Engeli, Rihoux, and Rothmayr Allison (2014: 100): ‘there is no rule that is set in stone for choosing one solution over the other. It is recommended to examine all three solutions (parsimonious, intermediate and complex) and to concentrate on the solution in which one has the most methodological confidence.’

Arguably, the elephant in the room here is neither a specific QCA solution type, nor the QCA method as such, but the prevalence of limited empirical diversity in the social world. Given this, it is all the more important to be well aware of the strengths and limitations of choosing a specific QCA approach for public policy analysis, and the implications of that choice on the interpretation of the results.

NOTES

1. The parts of this chapter describing the QCA technique and the types of approaches to QCA are excerpts from E. Thomann and M. Maggetti. 2017. Designing Research with Qualitative Comparative Analysis (QCA): Approaches, Challenges, Tools. *Sociological Methods & Research*, doi:10.1177/0049124117729700, copyright © 2017 by Sage Publications. Reprinted by permission of Sage Publications, Inc.
2. Insufficient but Non-redundant parts of Unnecessary but Sufficient (INUS) conditions (Mackie 1965; Mahoney, Kimball, and Koivu 2009).
3. Since this is an abstract example which does not lend itself to theoretical assumptions, I abstain from discussing the intermediate solution here.
4. PS M2 fails to do justice to M under any approach: not only does it fail to identify the causal relevance of A, it also grants causal relevance to condition C.
5. PS M2 uses row 5 as a simplifying assumption.

REFERENCES

- Adcock, R. and Collier, D. (2001). Measurement validity: a shared standard for qualitative and quantitative research. *American Political Science Review* 95 (3): 529–46.
- Baumgartner, F. R. and Jones, B. D. (1993). *Agendas and Instability in American Politics*. Chicago, IL: University of Chicago Press.
- Baumgartner, M. (2015). Parsimony and causality. *Quality & Quantity* 49 (2): 839–56.
- Baumgartner, M. and Thiem, A. (2017a). Model ambiguities in configurational comparative research. *Sociological Methods & Research* 46 (4): 954–87.
- Baumgartner, M. and Thiem, A. (2017b). Often trusted but never (properly) tested: evaluating qualitative comparative analysis. *Sociological Methods & Research*. doi:0049124117701487.
- Beach, D. and Rohlfing, I. (2018). Integrating cross-case analyses and process tracing in set-theoretic research strategies and parameters of debate. *Sociological Methods & Research* 47 (1): 3–36.
- Befani, B. (2013). Between complexity and generalization: addressing evaluation challenges with QCA. *Evaluation* 19 (3): 269–83.
- Befani, B. and Sager, F. (2006). QCA as a tool for realistic evaluations. In B. Rihoux and H. Grimm (Eds.), *Innovative Comparative Methods for Policy Analysis*. Boston, MA: Springer, 263–84.
- Berg-Schlosser, D. and De Meur, G. (2009). Comparative research design: case and variable selection. In B. Rihoux and C. C. Ragin (Eds.), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks, CA: Sage, 19–32.
- Berg-Schlosser, D., De Meur, G., Rihoux, B., and Ragin, C. C. (2009). Qualitative comparative analysis (QCA) as an approach. In B. Rihoux and C. C. Ragin (Eds.), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks, CA: Sage, 1–18.
- Blatter, J. and Blume, T. (2008). In search of co-variance, causal mechanisms or congruence? Towards a plural understanding of case studies. *Swiss Political Science Review* 14 (2): 315–356.
- Boon, J. and Verhoest, K. (2014). On a diet: explaining differences in overhead among public agencies in the era of austerity. *Public Performance & Management Review* 38 (2): 234–60.
- Börzel, T. and Risse, T. (2003). Conceptualizing the domestic impact of Europe. In K. Featherstone and C. Radaelli (Eds.), *The Politics of Europeanization*. Oxford: Oxford University Press, 57–80.

- Braumoeller, B. F. (2015). Guarding against false positives in qualitative comparative analysis. *Political Analysis* 23 (4): 471–87.
- Braun, D. and Gilardi, F. (2006). Taking ‘Galton’s problem’ seriously: towards a theory of policy diffusion. *Journal of Theoretical Politics* 18 (3): 298–322.
- Byrne, D. and Ragin, C. C. (Eds.) (2009). *The SAGE Handbook of Case-Based Methods*. Los Angeles, CA: Sage.
- Cacciatore, F., Natalini, A., and Wagemann, C. (2015). Clustered Europeanization and national reform programmes: a qualitative comparative analysis. *Journal of European Public Policy* 22 (8): 1186–211.
- Caren, N. and Panofsky, A. (2005). TQCA: a technique for adding temporality to qualitative comparative analysis. *Sociological Methods & Research* 34 (2): 147–72.
- Cooper, B. and Glaesser, J. (2016). Exploring the robustness of set theoretic findings from a large n fsQCA: an illustration from the sociology of education. *International Journal of Social Research Methodology* 19 (4): 445–59.
- Cronqvist, L. and Berg-Schlusser, D. (2009). Multi-value QCA (MVQCA). In B. Rihoux and C. C. Ragin (Eds.), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks, CA: Sage, 69–86.
- Damonte, A. (2014). Policy tools for green growth in the EU15: a qualitative comparative analysis. *Environmental Politics* 23 (1): 18–40.
- De Block, D. and Vis, B. (2019). Addressing the challenges related to transforming qualitative into quantitative data in qualitative comparative analysis. *Journal of Mixed Methods Research* 13 (4): 503–35.
- Dekker, R. and Scholten, P. (2017). Framing the immigration policy agenda: a qualitative comparative analysis of media effects on Dutch immigration policies. *The International Journal of Press/Politics* 22 (2): 202–22.
- Duşa, A. (2018). *QCA with R: A Comprehensive Resource*. New York: Springer.
- Eliason, S. R. and Stryker, R. (2009). Goodness-of-fit tests and descriptive measures in fuzzy-set analysis. *Sociological Methods & Research* 38 (1): 102–46.
- Elmore, R. F. (1985). Forward and backward mapping: reversible logic in the analysis of public policy. In K. Hanf and T. A. J. Toonen (Eds.), *Policy Implementation in Federal and Unitary Systems*. Dordrecht: Martinus Nijhoff, 33–70.
- Emmenegger, P. (2011). How good are your counterfactuals? Assessing quantitative macro-comparative welfare state research with qualitative criteria. *Journal of European Social Policy* 21 (4): 365–80.
- Emmenegger, P., Kvist, J., and Skaaning, S. (2013). Making the most of configurational comparative analysis: an assessment of QCA applications in comparative welfare-state research. *Political Research Quarterly* 66 (1): 185–90.
- Engeli, I., Rihoux, B., and Rothmayr Allison, C. (2014). Intermediate-N comparison: configurational comparative methods. In I. Engeli and C. Rothmayr Alison (Eds.), *Comparative Policy Studies: Conceptual and Methodological Challenges*. Basingstoke: Palgrave Macmillan, 85–108.
- Falletti, T. and Lynch, J. (2009). Context and causal mechanisms in political analysis. *Comparative Political Studies* 42 (9): 1143–66.
- Fischer, M. (2015). Institutions and coalitions in policy processes: a cross-sectoral comparison. *Journal of Public Policy* 35 (2): 245–68.
- Fischer, M. and Maggetti, M. (2017). Qualitative comparative analysis and the study of policy processes. *Journal of Comparative Policy Analysis: Research and Practice* 19 (4): 345–61.
- Fiss, P. C., Sharapov, D., and Cronqvist, K. (2013). Opposites attract? Opportunities and challenges for integrating large-n QCA and econometric analysis. *Political Research Quarterly* 66 (1): 191–7.
- Foschi, M. (1997). On scope conditions. *Small Group Research* 28 (4): 535–55.
- Funnell, S. and Rogers, P. (2011). *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco, CA: Jossey-Bass.
- Garcia-Castro, R. and Ariño, M. A. (2016). A general approach to panel data set-theoretic research. *Journal of Advances in Management Sciences & Information Systems* 2: 63–76.
- Gerring, J. (2004). What is a case study and what is it good for? *American Political Science Review* 98 (2): 341–54.
- Gerrits, L. M. and Verweij, S. (2016). Qualitative comparative analysis as a method for evaluating complex cases: an overview of literature and a stepwise guide with empirical application. *Zeitschrift für Evaluation* 15 (1): 7–22.

- Gerrits, L. M. and Verweij, S. (2018). *The Evaluation of Complex Infrastructure Projects: A Guide to Qualitative Comparative Analysis*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Goertz, G. (2006). Assessing the trivialness, relevance, and relative importance of necessary or sufficient conditions in social science. *Studies in Comparative International Development* 41 (2): 88–109.
- Goertz, G. and Mahoney, J. (2005). Two-level theories and fuzzy-set analysis. *Sociological Methods & Research* 33 (4): 497–538.
- Goertz, G. and Mahoney, J. (2006). Negative case selection: the possibility principle. In G. Goertz (Ed.), *Social Science Concepts: A User's Guide*. Princeton, NJ: Princeton University Press, 177–210.
- Goertz, G. and Starr, H. (2003). *Necessary Conditions: Theory, Methodology, and Applications*. Lanham, MD: Rowman & Littlefield.
- Greckhamer, T., Misangyi, V. F., and Fiss, P. C. (2013). The two QCAs: from a small-n to a large-n set theoretic approach. In P. C. Fiss, B. Cambré, and A. Marx (Eds.), *Configurational Theory and Methods in Organizational Research*. Bingley: Emerald Group, 49–75.
- Haesebrouck, T. (2015). The added value of multi-value qualitative comparative analysis. *Forum Qualitative Social Research/Sozialforschung* 17 (1): article 12.
- Hino, A. (2009). Time-series QCA. *Sociological Theory and Methods* 24 (2): 247–65.
- Hinterleitner, M., Sager, F., and Thomann, E. (2016). The politics of external approval: explaining the IMF's evaluation of austerity programs. *European Journal of Political Research* 55 (3): 549–67.
- Hooghe, L. and Marks, G. (2003). Unraveling the central state, but how? Types of multi-level governance. *American Political Science Review* 97 (2): 233–43.
- Hooijer, G. and Picot, G. (2015). European welfare states and migrant poverty: the institutional determinants of disadvantage. *Comparative Political Studies* 48 (14): 1879–904.
- Howlett, M., Ramesh, M., and Perl, A. (1995). *Studying Public Policy: Policy Cycles and Policy Subsystems*. Oxford: Oxford University Press.
- Jann, W. and Wegrich, K. (2007). Theories of the policy cycle. In F. Fischer and G. J. Miller (Eds.), *Handbook of Public Policy Analysis: Theory, Politics and Methods*. Boca Raton, FL: CRC Press, 43–62.
- Jano, D. (2016). Compliance with EU legislation in the pre-accession countries of South East Europe (2005–2011): a fuzzy-set qualitative comparative analysis. *Journal of European Integration* 38 (1): 1–22.
- Jones, M. and McBeth, M. (2010). A narrative policy framework: clear enough to be wrong? *Policy Studies Journal* 38 (2): 329–53.
- Kingdon, J. W. (1984). *Agendas, Alternatives, and Public Policies*. Boston, MA: Little, Brown.
- Knill, C. and Tosun, J. (2012). *Public Policy: A New Introduction*. Basingstoke: Palgrave Macmillan.
- Knoepfel, P., Larrue, C., Hill, M., and Varone, F. (2011). *Public Policy Analysis*. Bristol: Policy Press.
- Lipsky, M. (2010 [1980]). *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.
- Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly* 2 (4): 245–64.
- Maggetti, M. and Levi-Faur, D. (2013). Dealing with errors in QCA. *Political Research Quarterly* 66 (1): 198–204.
- Maggetti, M., Radaelli, C. M., and Gilardi, F. (2012). *Designing Research in the Social Sciences*. London: Sage.
- Mahoney, J. and Goertz, G. (2006). A tale of two cultures: contrasting quantitative and qualitative research. *Political Analysis* 14 (3): 227–49.
- Mahoney, J., Kimball, E., and Koivu, K. I. (2009). The logic of historical explanation in the social sciences. *Comparative Political Studies* 42 (1): 114–46.
- Mahoney, J. and Sweet Vanderpoel, R. (2015). Set diagrams and qualitative research. *Comparative Political Studies* 48 (1): 65–100.
- Matland, R. E. (1995). Synthesizing the implementation literature: the ambiguity-conflict model of policy implementation. *Journal of Public Administration Research and Theory* 5 (2): 145–74.
- Meuer, J., Rupietta, C., and Backes-Gellner, U. (2016). Layers of co-existing innovation systems. *Research Policy* 44 (4): 888–910.
- Mikkelsen, K. S. (2017). Negative case selection: justifications and consequences for set-theoretic MMR. *Sociological Methods & Research* 46 (4): 739–71.

- Nieto Morales, F., Wittek, R., and Heyse, L. (2015). Organizational pathways to compliant reform implementation: evidence from Mexican civil service reform. *Public Administration* 93 (3): 646–62.
- Oana, I.E., Schneider, C.Q. and E. Thomann. (2020). *Qualitative Comparative Analysis (QCA) using R: A Gentle Introduction*. Cambridge: Cambridge University Press.
- Ostrom, E. (2011). Background on the institutional analysis and development framework. *Policy Studies Journal* 39 (1): 7–27.
- Pahl-Wostl, C. and Knieper, C. (2014). The capacity of water governance to deal with the climate change adaptation challenge: using fuzzy set qualitative comparative analysis to distinguish between polycentric, fragmented and centralized regimes. *Global Environmental Change* 29 (1): 139–54.
- Patton, M. Q. (1997). *Utilization-Focused Evaluation*. Thousand Oaks, CA: Sage.
- Pattyn, V. and Brans, M. (2014). Explaining organizational variety in evaluation quality assurance: which conditions matter? *International Journal of Public Administration* 37 (6): 363–75.
- Pattyn, V., Molenveld, A., and Befani, B. (2019). Qualitative comparative analysis as an evaluation tool: lessons from an application in development cooperation. *American Journal of Evaluation* 40 (1): 55–74.
- Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. London: Sage.
- Peters, B. G. (1998). *Comparative Politics, Theory and Methods*. Basingstoke: Palgrave Macmillan.
- Pierson, P. (1993). When effect becomes cause: policy feedback and political change. *World Politics* 45 (4): 595–628.
- Radaelli, C. and De Francesco, F. (2010). Regulatory impact assessment. In R. Baldwin, M. Cave, and M. Lodge (Eds.), *The Oxford Handbook of Regulation*. Oxford: Oxford University Press, 279–301.
- Radaelli, C. M. and Wagemann, C. (2019). What did I leave out? Omitted variables in regression and qualitative comparative analysis. *European Political Science* 18 (2): 275–90.
- Ragin, C. C. (2014 [1987]). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley, CA: University of California Press.
- Ragin, C. C. (2000). *Fuzzy-Set Social Science*. Chicago, IL: University of Chicago Press.
- Ragin, C. C. (2008a). Measurement versus calibration: a set-theoretic approach. In J. M. Box-Steffensmeier, H. E. Brady, and D. Collier (Eds.), *The Oxford Handbook of Political Methodology*. Oxford: Oxford University Press, 174–98.
- Ragin, C. C. (2008b). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of Chicago Press.
- Ragin, C. C. and Fiss, P. C. (2017). *Intersectional Inequality: Race, Class, Test Scores, and Poverty*. Chicago, IL: University of Chicago Press.
- Rihoux, B. (2013). Qualitative comparative analysis (QCA), anno 2013: reframing the comparative method's seminal statements. *Swiss Political Science Review* 19 (2): 233–45.
- Rihoux, B., Alamos, P., Bol, D., Marx, A., and Rezsöházy, I. (2013). From niche to mainstream method? A comprehensive mapping of QCA applications in journal articles from 1984 to 2011. *Political Research Quarterly* 66 (1): 175–84.
- Rihoux, B. and Ragin, C. C. (Eds.) (2009). *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks, CA: Sage.
- Rihoux, B., Rezsöházy, I., and Bol, D. (2011). Qualitative comparative analysis (QCA) in public policy analysis: an extensive review. *German Policy Studies* 7 (3): 9–82.
- Rohlfing, I. (2018). Power and false negatives in qualitative comparative analysis: foundations, simulation and estimation for empirical studies. *Political Analysis* 26 (1): 72–89.
- Rohlfing, I. and Schneider, C. Q. (2013). Improving research on necessary conditions: formalized case selection for process tracing after QCA. *Political Research Quarterly* 66 (1): 220–235.
- Rohlfing, I. and Schneider, C. Q. (2016). A unifying framework for causal analysis in set-theoretic multimethod research. *Sociological Methods & Research* 47 (1): 37–43.
- Sabatier, P. A. (1988). An advocacy coalition framework of policy change and the role of policy-oriented learning therein. *Policy Sciences* 21 (2): 129–68.
- Sabatier, P. and Mazmanian, D. (1980). The implementation of public policy: a framework of analysis. *Policy Studies Journal* 8 (4): 538–60.
- Sabatier, P. and Weible, C. (Eds.) (2017). *Theories of the Policy Process*. Boulder, CO: Westview Press.
- Sager, F. and Andereggen, C. (2012). Dealing with complex causality in realist synthesis: the promise of qualitative comparative analysis. *American Journal of Evaluation* 33 (1): 60–78.

- Sager, F. and Thomann, E. (2017). Multiple streams in member state implementation: politics, problem construction and policy paths in Swiss asylum policy. *Journal of Public Policy* 37 (3): 287–314.
- Schneider, A. and Ingram, H. (1993). Social construction of target populations: implications for politics and policy. *American Political Science Review* 87 (2): 334–47.
- Schneider, C. Q. (2016). Real differences and overlooked similarities: set-methods in comparative perspective. *Comparative Political Studies* 49 (6): 781–92.
- Schneider, C. Q. (2018). Realists and idealists in QCA. *Political Analysis* 26 (2): 246–54.
- Schneider, C. Q. and Rohlfing, I. (2013). Combining QCA and process tracing in set-theoretic multi-method research. *Sociological Methods & Research* 42 (4): 559–97.
- Schneider, C. Q. and Rohlfing, I. (2016). Case studies nested in fuzzy-set QCA on sufficiency: formalizing case selection and causal inference. *Sociological Methods & Research* 45 (3): 526–68.
- Schneider, C. Q. and Wagemann, C. (2010). Standards of good practice in qualitative comparative analysis (QCA) and fuzzy sets. *Comparative Sociology* 9 (3): 397–418.
- Schneider, C. Q. and Wagemann, C. (2012). *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. New York: Cambridge University Press.
- Schneider, C. Q. and Wagemann, C. (2013). Doing justice to logical remainders in QCA: moving beyond the standard analysis. *Political Research Quarterly* 66 (1): 211–20.
- Schneider, C. Q. and Wagemann, C. (2016). Assessing ESA on what it is designed for: a reply to Cooper and Glaesser. *Field Methods* 28 (3): 316–21.
- Seawright, J. and Collier, D. (2010). Glossary. In H. E. Brady and D. Collier (Eds.), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield, 313–60.
- Shahidi, F. V. (2015). Welfare capitalism in crisis: a qualitative comparative analysis of labour market policy responses to the great recession. *Journal of Social Policy* 44 (4): 659–86.
- Skaaning, S. (2011). Assessing the robustness of crisp-set and fuzzy-set QCA results. *Sociological Methods & Research* 40 (2): 391–408.
- Tanner, S. (2014). QCA and causal inference: a poor match for public policy research. *Qualitative & Multi-Method Research* 12 (1): 15–24.
- Thiem, A. (2014). Unifying configurational comparative methods: generalized-set qualitative comparative analysis. *Sociological Methods & Research* 43 (2), 313–37.
- Thiem, A. (2016a). Conducting configurational comparative research with qualitative comparative analysis: a hands-on tutorial for applied evaluation scholars and practitioners. *American Journal of Evaluation* 38 (3): 420–433.
- Thiem, A. and Baumgartner, M. (2016a). Back to square one: a reply to Munck, Paine, and Schneider. *Comparative Political Studies* 49 (6): 801–6.
- Thiem, A., Baumgartner, M., and Bol, D. (2016). Still lost in translation! A correction of three misunderstandings between configurational comparativists and regression analysts. *Comparative Political Studies* 49 (6): 742–74.
- Thomann, E. (2015). Customizing Europe: transposition as bottom-up implementation. *Journal of European Public Policy* 22 (10): 1368–87.
- Thomann, E. (2018). *Customized Implementation of European Union Food Safety Policy: United in Diversity?* Cham: Palgrave Macmillan.
- Thomann, E. (2019). Qualitative comparative analysis (QCA) as a tool for street-level bureaucracy research. In P. Hupe (Ed.), *Research Handbook on Street-Level Bureaucracy: The Ground Floor of Government in Context*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing, 370–391.
- Thomann, E. and Maggetti, M. (2017). Designing research with qualitative comparative analysis (QCA): approaches, challenges, tools. *Sociological Methods & Research*. doi:10.1177/0049124117729700.
- Thomann, E., van Engen, N., and Tummers, L. (2018). The necessity of discretion: a behavioral evaluation of bottom-up implementation theory. *Journal of Public Administration Research and Theory* 28 (4): 583–601.
- Toshkov, D. (2016). *Research Design in Political Science*. London: Palgrave Macmillan.
- True, J. L., Jones, B., and Baumgartner, F. R. (1999). Punctuated-equilibrium theory: explaining stability and change in American policymaking. In A. Sabatier (Ed.), *Theories of the Policy Process*. Boulder, CO: Westview Press, 97–115.

- van der Heijden, J. (2015). The role of government in voluntary environmental programmes: a fuzzy set qualitative comparative analysis. *Public Administration* 93 (3): 576–92.
- Varone, F., Rihoux, B., and Marx, A. (2006). A new method for policy evaluation? In B. Rihoux and H. Grimm (Eds.), *Innovative Comparative Methods for Policy Analysis*. New York: Springer, 213–36.
- Wagemann, C., Buche, J., and Siewert, M. B. (2016). QCA and business research: work in progress or a consolidated agenda? *Journal of Business Research* 69 (7): 2531–40.
- Williams, T. and Gemperle, S. M. (2017). Sequence will tell! Integrating temporality into set-theoretic multi-method research combining comparative process tracing and qualitative comparative analysis. *International Journal of Social Research Methodology* 20 (2): 121–35.
- Winter, S. C. (2003). Implementation perspectives: status and reconsideration. In B. G. Peters and J. Pierre (Eds.), *Handbook of Public Administration*. London: Sage, 212–22.
- Zhang, J. (2017). On the minimization principle in the Boolean approach to causal discovery. In S. C. Yang, K. Y. Lee, and H. Ono (Eds.), *Philosophical Logic: Current Trends in Asia*. Singapore: Springer, 79–94.

16. Process tracing for comparative policy analysis: a realist approach

Guillaume Fontaine

1. INTRODUCTION: A POPULAR METHOD

1.1 The Milestones of Process Tracing in Social Sciences

Process tracing has raised increasing interest for the past decade, thus enriching the toolbox of comparative policy analysis together with historical analysis and qualitative comparative analysis (QCA). Although process tracing methods have been used in health sciences since the nineteenth century, for instance in John Snow's research on the causes of cholera epidemic (Snow 1855), they only raised the attention of social scientists in the late twentieth century in history and international relations (George and McKeown 1985; Skocpol and Somers 1980). As of mid-2018 a review of the subject "process tracing" or "causal mechanism" combined with "policy" produced 538 records in the Web of Science, 95% of which had been published after 2005, mostly in political science, international relations, public administration, and economics. When adding the subject "path dependence", a theory generally related with process tracing in political economy and historical sociology (Bennett and Elman 2006), the result reached 1,143 records, 93% of which had been published since 2005.

The publication of *Case Studies and Theory Development in the Social Sciences* (George and Bennett 2005) was a milestone, judging by its popularity on social networks.¹ In their plea for rigorous case study, Alexander George and Andrew Bennett went beyond the never-ending and somewhat reiterative debate on quantitative versus qualitative methods (see Blatter and Haverland 2012; Brady and Collier 2010; Mahoney and Goertz 2006) and upgraded the literature by addressing the problems of transparency, replicability, and small-N comparison in social sciences. In their own words, process tracing is but one type of case study method, but it is the most trustworthy when it comes to establish a causal relationship within a single case, for it provides a template for evidence assessment in any ongoing research linking a trigger or a cause X to an outcome Y.

Yet given its increasing popularity among scholars from many different backgrounds, process tracing was likely to suffer from conceptual stretching and lose its analytical precision. Hence the challenge was to reinforce its philosophical and methodological grounds. The textbook *Process Tracing Methods: Foundations and Guidelines* (Beach and Pedersen 2013) made a decisive contribution to this enterprise by presenting a step-by-step recipe of process tracing applications. One major contribution of this book was to identify three types of process tracing and explain their practical implications. On one hand, theory-driven process tracing can be used either for theory-building (in order to theorize a portable causal mechanism) or theory-testing (for a confirming or disconfirming purpose); on the other hand, casuistic process tracing can be used to explain a *sui generis* outcome in a unique or an exceptional context (such as a revolution or an international crisis). The authors also placed causal mecha-

nisms at the core of process tracing, and provided a template to design and perform empirical tests, to gather and interpret empirical evidence, and to select relevant cases.

Soon afterwards Jeffrey Checkel and Andrew Bennett invited a group of scholars to share their experience in *Process Tracing: From Metaphor to Analytic Tool* (Bennett and Checkel 2015a). Although their case studies were still mainly related to international relations, as in former publications by George and Bennett, the editors drew some conclusions and made meaningful recommendations for other disciplines and research areas. In particular, they presented a list of ten good practices to enlighten scholars engaging in process tracing. Most strikingly, they raised the bar for further research by bringing in the Bayesian logic of likelihood to the discussion, in an attempt to provide a way of assessing evidence through formal process tracing. One of the most challenging requirements in process tracing is to confer an equanimous treatment to the main hypothesis and all alternate hypotheses. This can be done in different ways, using congruence analysis, counterfactual reasoning, or experiments, all of which are time-consuming and raise technical problems regarding case selection, data recollection and interpretation, and so forth. The inclusion of Bayesian statistics into the process tracing apparel offered a way to deal with this problem at a minor cost–benefit ratio.

1.2 Process Tracing Applications to Policy Analysis

Although it was originally designed for within-case studies, process tracing is now commonly used to conduct small-N comparison (Beach and Pedersen 2016a), and combined with QCA in a multi-method research design (Beach 2017, 2018; Rohlfing and Schneider 2016). Further, there is a growing consent about its potential contribution to comparative policy analysis, to advance the theory of change in policy evaluation (Befani and Mayne 2014; Befani and Stedman-Bryce 2016), to assess the policy process in a transparent and systematic way (Kay and Baker 2015), or to improve the decision-making process in public administration (Charbonneau et al. 2016).

In comparative politics, the relatively new focus on processes may be a way to fill in the gap between scholars' ontology and their methods, which appeared while evolving from an independent variable conception of causation toward a configurational one (Hall 2003: 387). This way it became a complement to address problems faced by probabilistic methods, such as multi-causality, equifinality, and causal heterogeneity (Blatter and Haverland 2014: 64). Yet many different techniques and uses of process tracing coexist (even within policy studies), according to the kind of process being traced (Trampusch and Palier 2016), so that pretending to achieve a synthesis seems preposterous. Instead, this chapter advocates for a realist approach to process tracing, aimed at explaining the causal mechanism linking the adoption of a policy aim to a policy outcome, beyond directly observable phenomena (Fontaine, Fuentes, and Narváez 2018; Fontaine, Narváez, and Paz 2017).

Section 2 presents a solution to the problem of alignment between ontology and methods with realist process tracing. This method is philosophically grounded in scientific realism, according to which causal explanations of the world are possible beyond non-observable but detectable forces. It aims at opening the black box of causation through deep within-case studies and small-N comparison, for which it focuses on the causal mechanism that explains how an event X triggers a process leading to an outcome Y. Section 3 offers a template to conduct research with this method, based on a five-step protocol including: (1) the theorization and operationalization of a causal mechanism; (2) the selection of a typical case on this mech-

anism for theory-building; (3) the design of empirical tests based on the expected observations for each part of the causal process; (4) evidence gathering and assessment through Bayesian analysis; and (5) cross-case comparison with another typical or a deviant case for a confirming or a disconfirming purpose. The chapter concludes with a reflection on how to improve the transparency and replicability of this framework for further research.

2. METHODOLOGY

The realist approach to process tracing is grounded in scientific realism, a methodology combining a dualist ontology with a transfactual conception of knowledge. It aims at opening the black box of causation through deep within-case studies and small-N comparison. It focuses on the causal mechanism that explains how a factor or an event X triggers a process leading to an outcome Y.

2.1 Aligning Process Tracing with Realism

Realism is often presented as a third way in the neo-positivism versus constructivism debate in social sciences, involving philosophical, epistemological, theoretical, and methodological issues (Marsh and Furlong 2010; Wendt 1999). In the present chapter it is understood as a methodology that combines a philosophical ontology (i.e. a conception of the world) and a scientific ontology (i.e. a conception of the knowledge of the world) (Jackson 2016; Sayer 2010). Such a distinction goes against the “epistemic fallacy” (Bhaskar 2008 [1978]: 26) according to which a statement about the world can be reduced to a statement about knowledge. This definition has two important consequences for the following discussion. First, it implies that there is no reason to conflate methods and techniques with methodology (as in Hall 2003). Second, it argues against a deterministic relationship between ontology, epistemology, and methods (as in Marsh and Furlong 2010).

Realism is a combination of dualism, a philosophical ontology that states a separation between the mind and the world, and transfactualism, a scientific ontology that reckons the existence of non-observable but detectable factors of causation. This definition is based on Roy Bhaskar’s “transcendental realism” (Bhaskar 2008 [1978]) which was also coined as “critical realism” (Archer et al. 1998; Jackson 2016; Sayer 2010), as opposed to Kant’s “transcendental idealism” and Hume’s “empirical realism”. On the one hand, Bhaskar argued against Kant that the world cannot be reduced to a construction of the human mind, so there is actually a world outside the mind, which defines a mind–world gap that needs to be bridged. On the other hand he argued against Hume that knowledge about the world cannot be reduced to atomistic events, so this world is an open system where regularities are more the exception than the rule. Scientific knowledge is a social construction whose objects are structures and mechanisms that generate phenomena (Bhaskar 2008 [1978]: 15).

This methodology departs from the production of law-like theories implied by the neo-positivist conception of causation (Hempel and Oppenheim 1948), and from the logic of falsification (Popper 2014 [1959]) because non-directly observable and contextual factors cannot be treated in the same way as directly observable phenomena. It acknowledges the value of qualitative methods per se for causal explanations, regardless of their external validity. This does not mean realism commands the use of qualitative methods but, conversely, that

qualitative methods such as congruence analysis or process tracing aimed at producing causal explanations are logically aligned with realism (Morgan 2016; Sayer 2000).

Realist scholars seek to identify patterns of causation rather than building models of predictive explanations (Bhaskar 2008 [1978]). They deal with causal forces through a transfactual analysis, as opposed to the phenomenological analysis conducted by neo-positivists. By extending the limits of knowledge to a priori non-observable phenomena (Jackson 2016: 56), they produce non-predictive explanations of the world through laboratory investigations or transcendental arguments. They conduct contrasting case-oriented comparisons as a way of elucidating causal powers. Therefore they are most compelling when engaging in small-N comparison and deep accounts of typical, exemplary, or influential case studies. In the absence of a laboratory to conduct experiments on a policy design, fuzzy-set QCA and process tracing may serve to eliminate non-explicative factors and to identify causal mechanisms as combinations of factors explaining a particular outcome (Beach and Pedersen 2016a).

In a nutshell aligning process tracing with realism means conceiving it as both explanatory and transfactual. This has important consequences for a set-theoretic research design, as described below.

2.2 Realist Process Tracing as a Set-Theoretic Method

Realist process tracing fits in a deterministic logic of causation, as opposed to the probabilistic logic of most neo-positivist methods. This neither precludes the combination of qualitative and quantitative techniques, nor does it preempt theoretical debates between agent-based and institutional frameworks.

Cross-case comparison is the only way to provide a case study for some degree of generalization or external validity (Peters 2013), but this means two different things for neo-positivists and realists. For neo-positivists the external validity of a causal relationship depends on the co-variance between independent and dependent variables (Seawright 2016). Most scholars consider qualitative methods as irrelevant to establish or test causality, hence using them as a complement to statistics in so-called “mixed methods” (Seawright 2016; Seawright and Gerring 2008). To them process tracing can only be used as a subordinate method to shed light on a particular aspect of the theory. Even those who value case studies as much as quantitative analysis use them in search of sufficient and necessary conditions (Collier 2011; Goertz and Starr 2003; Mahoney 2001), which are law-like concepts that eventually require the kind of cross-case comparison which is guided by a probabilistic conception of causality.

For realists, the possibility to draw some generalizations from case studies depends on the evidence that these variables are actually connected by entities at work within a context (Beach and Pedersen 2016a). Therefore cross-case comparison has a different purpose. Instead of seeking for a probabilistic relationship between independent and dependent variables, it aims at confirming or disconfirming a deterministic relationship. Assessing the co-variance between X and Y requires multiplying observations to calculate the probability that a variation in X causes a variation in Y or that in the absence of a variation in X there is no variation in Y. Assessing a causal process between X and Y implies looking within a case at the way a series of entities (organizations, actors, events, etc.) interact together and with the context, so that a variation in X triggers a process leading to an outcome Y. This does not mean that both logics are conflicting – as a matter of fact their complementarity is at the heart of multi-method research design (Lorentzen, Fravel, and Paine 2016). But they support techniques that need

to be applied separately and for different purposes, which implies a careful methodological alignment.

Further, if realist process tracing consists in analyzing causal forces in order to make a statement about the relationship between a trigger X and an outcome Y, then it is necessary to consider both the trigger and the outcome as parts of the process. Consequently there is no great interest in conducting an “intensive” process tracing (Falleti 2016) that would focus only on what happens after a trigger has been activated or before an outcome has been produced. Excluding the trigger or the outcome from a case study can only mean there is a process within a process. This may produce insightful knowledge on the entities and activities under scrutiny but it does not allow drawing any conclusion beyond speculation on a causal process as a whole. This is a major limitation of the use of path-dependent explanations in policy studies, in spite of the irreplaceable contribution of historical institutionalism to the analysis of causal complexity through long-term framing and detailed analysis of rare events (Bennett and Elman 2006). Based on long-term process tracing, this theory excels at explaining policy stability after critical junctures leading to a third-order or paradigmatic policy change (Hall 1993; Kern, Kuzemko, and Mitchell 2014). But it fails to produce explanations of change due to policy learning over a short or a medium term (or first- and second-order change) (Howlett and Rayner 2006; Kay 2005).

Eventually, process tracing is about methods, not about theory. Therefore there is no reason to reduce it to an agent-based model (Hedström and Swedberg 1998; Hedström and Ylikoski 2010). Agent-based modeling may be effective to explain decision making, while being of limited value to provide a full-range explanation of the complex causality at stake in an implementation gap or a policy change. Realist process tracing can deal with two different kinds of processes: realized and anticipated. To give but two examples, while the former may involve multi-level interactions, like the international and the national context of fiscal policies in the European Union, the latter may refer to expected interactions or patterns of behavior, such as those predicted by the democratic peace theory (Rohlfing 2013). So a realist process tracing could both apply to an institutional explanation of a realized process and to an agent-based explanation of an anticipated process.

In a nutshell, realist process tracing follows a deterministic logic of causation and focuses on all the elements of a causal process, including the trigger, the outcome, and the causal mechanism in between, regardless of the theoretical argument made. It is therefore mandatory to align the definition of causal mechanisms with realism, as below.

2.3 A Realist View on Causal Mechanisms

According to their background, scholars have theorized causal mechanisms alternatively as parts of a narrative (Crasnow 2017; Ruback 2010) or an analytical process (Beach 2013), as a theoretical (Paquet and Broschek 2017), an ontological (Gerring 2007b), or a methodological issue (Mahoney 2016). Like process tracing, a causal mechanism is a polysemic concept whose definition can vary according to the phenomenon being observed (be it a historical event, a political decision, or a diplomatic negotiation) and to scholars’ philosophical and scientific ontology. In its broadest sense it describes a causal relationship linking an *explanans* X to an *explanandum* Y (Gerring 2007b: 166). In a narrower sense it is but one modality of causation among others, such as temporal and spatial contiguity, causal chains and process dynamics, causal combinations or configurations (Blatter and Haverland 2012: 91–2).

Causal mechanisms lie at the core of process tracing, which might otherwise consist in a descriptive exercise to account for a pathway or to build up a narrative on a series of events (Beach 2016). Realist process tracing aims at providing evidence of a causal mechanism that can complement other methods, in order to assess the causal relationship between X and Y. This kind of evidence is “mechanistic” but it can consist in both qualitative and quantitative data. As the “epistemic theory of causality” in health sciences (Russo and Williamson 2007) realism reckons the difference in kind between probabilistic and mechanistic evidence, and utilizes them both to produce a satisfactory explanation of a causal relationship between a trigger X and an outcome Y. To be sure, the difference lies between two kinds of evidence, not between two kinds of evidence-gathering methods (McKay Illari 2011).

These differences also come from different conceptions of causation. For neo-positivists, “social mechanisms” are part of an ongoing process of theory-building and testing, they provide “useful pieces of a theory” (Stinchcombe 1991: 367), partial explanations or intermediate paths to advance the grand theory (Hall 2013). From this standpoint, a causal mechanism is a set of intervening variables (King, Keohane, and Verba 1994), a black box that remains unstructured and within which there is a priori no hierarchy. Therefore “opening the black box” through process tracing consists in testing a theory based on a probabilistic model, by explaining either how the cogs and wheels work in a typical case or how a disruptive factor works in a deviant case (Brady, Collier, and Seawright 2010; Falleti and Lynch 2009). In any case, the higher the external validity, the better for the theory since the more generalizable the conclusions.

For realists, mechanisms are “elementary building blocks of middle-range theories”, which differ from law-like theories inasmuch as they are “unobserved analytical constructs” playing out to articulate the micro- and the macro-levels of social action (Hedström and Swedberg 1998: 6, 13). From that perspective, a causal mechanism is a set of causal process observations (George and Bennett 2005), a black box containing expected empirical observations that need to be turned into evidence of the cause of an outcome (Bennett 2015; Bennett and Checkel 2015b). Therefore, “opening the black box” consists in revealing a causal mechanism’s structure, hierarchy, and sequencing (Hedström and Ylikoski 2010): the higher the internal validity, the better for the theory although the conclusion may be hardly generalizable.

Whether the process is political, economic, social, or even psychological, whether the agent of causation is individual or collective, whether the time frame refers to long-lasting or instantaneous, unique or regular events, this does not change the fact that all entities or parts of a causal mechanism are effectively related in a way that links the action of a trigger to the production of an outcome. These entities are not mere elements that stick together by the art of magic; they enter into activities related to endogenous and exogenous factors (Machamer 2004). From that perspective, rather than a set of intervening variables, a causal mechanism is best described as a system of interacting parts. Another way of putting it is to say a causal mechanism is a combination of Insufficient but Necessary parts of an Unnecessary but Sufficient condition (Mackie 1965), an INUS complex, as long as the research focuses both on the substance (the identification of the entities that make up a mechanism) and the process (the identification of the activities these entities are involved in).

Eventually separating a mechanism from its context makes little sense for realist process tracing (but see Falleti and Lynch 2009). Since the method consists precisely in integrating the piecemeal entities and the contingency into a single causal mechanism, it is neither a descriptive exercise of how a theoretical mechanism should work all other things equal, nor a pre-

scriptive one to convert a case study into an ideal-type or a model. It is about explaining how a set of entities interact with a social context – how they work in context – or what contextual factors prevent them from producing an expected effect (Astbury and Leeuw 2010).

In a nutshell, the realist approach to process tracing conceives a causal mechanism as a combination of entities entering in activity to produce an outcome in a particular context.

The combination of dualism and transfactualism defines realism as a methodology according to which causal explanations of the world as an open system are self-sufficient middle-range theories. Regardless of the theoretical background supporting a research hypothesis, this methodology focuses on how a causal mechanism linking a trigger to an outcome sets a combination of entities in motion, through a detectable but non-directly observable causal force. The following section explains how to conduct such research, based on a policy design framework.

3. A TEMPLATE FOR REALIST PROCESS TRACING

This section presents a five-step research design including the process theorization and operationalization, empirical tests design, case selection, congruence analysis, and deep within-case study.

3.1 Step 1: Process Theorization

The problem of causal mechanism theorization can be summarized as follows: we believe (with a certain degree of confidence) that an event triggers a process leading to an outcome, but we don't know why. Hence realist process tracing is not only a method to gather evidence for a within-case study (Gerring 2007b), it is part of a set-theoretic multi-method research design (Rohlfing and Schneider 2016). The very fact that we believe or know something about a causal relationship even before engaging in process tracing implies that the research aims at building or testing a theoretical causal mechanism and that, at least, one other method has been utilized to formulate our hypothesis.

In the present example, the theoretical causal mechanism is based on a policy design framework (Fontaine, Medrano, and Narváez 2020). Policy design is first and foremost the result of continuous changes (either incremental or revolutionary) in the rules of governance and the proliferation of complex and wicked policy problems (Peters 2018). To address these problems, the combination of policy instruments as detectors and effectors for each policy area produces a virtually infinite number of policy mixes (Hood and Margetts 2007: 147). Yet when designing policies, governments do not start from *tabula rasa*, they would rather adapt to the changing context by altering a policy's aims and means (Howlett 2009).

In this framework the process to be traced goes from the adoption of a policy aim to the production of a policy outcome. Table 16.1 presents this causal process, each part of which is related to a main hypothesis.

The theoretical process rolls out as follows. The agenda-setting (T) triggers a three-part causal mechanism (A, B, C) producing a policy outcome (O).² The causal mechanism includes the formulation of a policy (A), and two outputs which are common to any public policy: cross-sectorial coordination (B) and political interplays (C). Confirming this mechanism would require observing new policy aims claimed by the government (T), a change in

Table 16.1 *The theoretical process of policy design*

Theoretical mechanism	Agenda-setting (T)	Policy formulation (A)	Cross-sectorial coordination (B)	Political interplays (C)	Policy outcome (O)
Operationalization	New policy aim claimed by government	Change in a sectorial policy	Change in the institutional system	Implementation style in state–society relationships	Public account of a policy outcome
Main hypothesis	Government aims at economic development through resource nationalism (H _T)	Government formulates a new oil policy based on resource nationalism (H _A)	Government centralizes institutional system to support new oil policy (H _B)	Government adopts a hierarchical implementation style to cope with non-state actors contest (H _C)	Government limits social control and citizens’ participation in the oil policy (H _O)
Alternate hypotheses	Government does not aim at securing development through resource nationalism (¬H _T)	Government does not formulate a new oil policy based on resource nationalism (¬H _A)	Government does not adapt institutional system to support new oil policy (¬H _B)	Government does not adopt a hierarchical implementation style to cope with non-state contest (¬H _C)	Government does not limit social control and participation in oil policy (¬H _O)

Source: Fontaine, Medrano, and Narváez (2020).

a policy area (A) and the institutional system (B), an implementation style in state–society relationships (C), and a public account of the policy outcome (O). Theoretically these parts are causally related (H=T:A:B:C:O) and the whole process is the sum of the trigger, the outcome, and the mechanism (H=H_T*H_A*H_B*H_C*H_O). This means the confirmation of the theoretical process depends on the confirmation of its weakest part (Beach and Pedersen 2013). Any potential loop effect or back and forth movement affecting the same policy is a process within the process, hence constituting itself a case study.

For example, our main hypothesis states that when a government aims at securing development through resource nationalism (H_T), this causes the formulation of a new oil policy based on resource nationalism (H_A). Consequently the government centralizes the existing institutional system to support the new oil policy (H_B), then adopts a hierarchical implementation style to cope with the contest by non-state actors (H_C), which ends in limiting social control and participation in the oil policy (H_O). Any other hypothesis is treated as a single alternate hypothesis (¬H) for each part of the mechanism. These alternate hypotheses state that a government does not aim at securing development through resource nationalism (¬H_T), neither do they formulate a new oil policy based on resource nationalism (¬H_A), nor do they adapt the institutional system to support the new oil policy (¬H_B). The government does not adopt a hierarchical implementation style to cope with the contest by non-state actors (¬H_C), and they do not limit social control and participation in the oil policy (¬H_O).

Unlike the main hypothesis, the alternate hypotheses are not bounded together by a causal relationship. They are “empirical counterfactual” elements (Rohlfing 2014) for each individual entity of the mechanism, the trigger, and the outcome. Further, an alternate hypothesis is not a competing hypothesis against which the causal mechanism should be tested. It just conflates all other possible hypotheses that may disconfirm the theoretical causal mechanism. In other words, only the main hypothesis (H) can be confirmed. The logic here is comparable to that of

a court trial, where a defendant may leave the court either guilty of charge or innocent, without the judge saying anything about other possible suspects. If it was to be disconfirmed, the main hypothesis should be discarded or reformulated but the alternate hypothesis would remain indeterminate (Beach and Pedersen 2013; Bennett and Checkel 2015b). This goes against the experimental logic of neo-positivist process tracing, where each possible hypothesis is expected to be treated with the same prior degree of confidence before being tested against available evidence (Fairfield and Charman 2017; Humphreys and Jacobs 2015).

Once the process has been theorized, the research can proceed with the design of empirical tests aimed at assessing the theoretical causal mechanism, the trigger, and the outcome.

3.2 Step 2: Empirical Tests Design

The design of empirical tests includes the identification of expected empirical observations if the theoretical mechanism stands, and the assessment of their probatory value if they result in a true positive result.

An expected empirical observation refers to what kind of evidence is needed to confirm the main hypothesis, so that the result would be disconfirming if no evidence was to be found or if evidence of the alternate hypothesis was to be found. The combination of instruments favored by a government is a clear-cut indicator of the nature and orientation of a policy design (Hood 1986), which provides a reliable source of expected empirical observations. The rationale behind this idea is that policy instruments constitute an INUS condition for a policy to actually exist (regardless of its inputs, outputs, and outcomes). Of all instruments classifications, the most parsimonious and widespread in comparative policy analysis rests upon the state's resources of nodality, authority, treasure, and organization (Hood 2007; Howlett 2011; Linder and Peters 1998; Ringeling 2005). Nodality refers to the situation of the government amidst information and social networks; authority refers to the official power and the production of legal norms by the state; treasure refers to the available financial resources and the power to emit money; organization refers to the state administration's capacity according to their human and physical resources (Hood 1986: 5–6).

Based on this taxonomy, Table 16.2 presents a typology of the expected empirical observations for each part of the causal process of policy design, if our main hypothesis stands.

If H_T stands, explicit claims of the government aiming at securing development through resource nationalism should be observed in the government's program (P_T1), in the constitutive regulation (P_T2), in the development model (P_T3), and in the design of the state apparatus (P_T4). If H_A stands, a change in the oil policy based on resource nationalism should be seen in the sectorial instruments of planning (P_A1), regulation (P_A2), budget allocation (P_A3), and public administration (P_A4). Likewise, if H_B stands, a centralization of the institutional system aimed at supporting the new policy should be observed in the cross-sectorial instruments of planning (P_B1), regulation (P_B2), budget allocation (P_B3), and public administration (P_B4). If H_C stands, the hierarchical implementation style to cope with non-state actors contest should be observed in the incumbent's assessments (P_C1), in the legislative process (P_C2), in the state's budget execution (P_C3), and in the local community administration (P_C4). If H_O stands, a public account of the policy outcome limiting social control and citizens' participation in the oil policy should appear in the limitations regarding access to the information system (P_O1), respect of the due process by the state agencies (P_O2), financial assessments by the government (P_O3), and autonomy by balance and control agencies (P_O4).

Table 16.2 *Expected empirical observations in a policy design process*

Hypothesis	Government aims at economic development through resource nationalism (H ₁)	Government formulates a new oil policy based on resource nationalism (H _A)	Government adapts the institutional system to support the new oil policy (H _B)	Government adopts a hierarchical implementation style to cope with non-state actors contest (H _C)	Government limits social control and citizens' participation in the oil policy (H _O)
Nodality instruments	Explicit claims in the government's program (P ₁ 1)	Change in the policy area planning (P _A 1)	Change in the cross-sectorial planning (P _B 1)	Traces in the incumbent's statements (P _C 1)	Limited access to the information system (P _O 1)
Authority instruments	Explicit claims by government in the constitutive regulation (P ₁ 2)	Change in the policy area regulation (P _A 2)	Change in the cross-sectorial regulation (P _B 2)	Traces in the legislative process (P _C 2)	Limited respect of the due process by state agencies (P _O 2)
Treasure instruments	Explicit claims by government in the development model (P ₁ 3)	Change in the policy area budget allocation (P _A 3)	Change in the cross-sectorial budget allocation (P _B 3)	Traces in the budget execution (P _C 3)	Limited financial assessments by the government (P _O 3)
Organization instruments	Explicit claims by government in the state apparatus design (P ₁ 4)	Change in the policy area administration (P _A 4)	Change in the cross-sectorial administration (P _B 4)	Traces in the local community administration (P _C 4)	Limited autonomy by balance and control agencies (P _O 4)

Source: Fontaine, Medrano, and Narváez (2020).

The confirming or disconfirming value of these observations depends simultaneously on their degree of certainty (the probability to find the evidence e , given the main hypothesis h , $p(e|h)$) and their uniqueness (the probability to find the evidence e , given the alternate hypothesis $\neg h$, $p(e|\neg h)$). The combination of certainty and uniqueness defines four types of empirical tests: straw-in-the-wind, smoking gun, hoop, and doubly-decisive (Collier 2011; Van Evera 1997). A straw-in-the-wind test is neither certain nor unique; a doubly-decisive test is both certain and unique; a smoking-gun test is not certain but unique; and a hoop test is certain but not unique.

A parsimonious way to assess formally this probatory value is to calculate the posterior confidence given the collected evidence $p(h|e)$ with Bayes likelihood theorem (Bennett 2006, 2015). Our prior confidence $p(h)$ depends on the previous knowledge supporting h , which can rely on an experiment, a formal model based on a standard linear regression, the results of former empirical tests, or even a subjective perception measured by opinion surveys. In absence of this bottom line, a random value of the prior confidence may be that of a flipping coin [$p(h) = 0.5$].

The posterior confidence can be calculated as follows:

$$p(h|e) = \frac{p(h)p(e|h)}{p(h)p(e|h)+p(\neg h)p(e|\neg h)} \text{ in a positive test;}$$

and

$$p(h|\neg e) = \frac{p(h)p(\neg e|h)}{p(h)p(\neg e|h)+p(\neg h)p(\neg e|\neg h)} \text{ in a negative test;}$$

Table 16.3 Bayesian formalization of conventional tests

Tests	p(h)	p (¬h)	p (e h)	p (¬e h)	p (e ¬h)	p (¬e ¬h)	p (h e)	p (h ¬e)	C1=p (h e)-p(h)	C2=p (h ¬e)-p(h)
Straw-in-the-wind	0.50	0.50	0.40	0.60	0.30	0.70	0.57	0.46	0.07	-0.04
Hoop	0.50	0.50	0.90	0.10	0.30	0.70	0.75	0.13	0.25	-0.38
Smoking gun	0.50	0.50	0.40	0.60	0.10	0.90	0.80	0.40	0.30	-0.10
Doubly-decisive	0.50	0.50	0.90	0.10	0.10	0.90	0.90	0.10	0.40	-0.40

Source: Fontaine, Medrano, and Narváez (2020).

where *h* is the hypothesis, *e* is the evidence, ¬*h* is the alternate hypothesis and ¬*e* is the absence of evidence (adapted from Bennett 2015).

Table 16.3 presents the results of positive and negative tests based on $[p(h) = 0.5]$.³

Straw-in-the-wind tests exhibit the lowest confirming (0.07) and disconfirming (-0.04) values. Conversely, doubly-decisive tests show the highest confirming (0.4) and disconfirming (-0.4) values. At an intermediate level, hoop tests show a lower confirming value than smoking gun tests (0.25 vs 0.3) but a higher disconfirming value (-0.38 vs. -0.1).

Once the expected empirical observations have been defined, the research can proceed to case selection.

3.3 Step 3: Case Selection

In a realist process tracing, case selection is neither based on the trigger nor on the outcome, but on the theoretical relationship between the one and the other (Beach and Pedersen 2016b; Beach and Rohlfing 2015). As a consequence, it is not affected by the dilemma of causes-of-effects versus effects-of-causes criteria (Goertz and Mahoney 2012) or by the problem of statistical bias raised by the selection on the dependent variable (Geddes 2003). Yet it is not straightforward and it requires at least two preliminary steps. First it implies a definition of the attributes of the trigger and the outcome, which corresponds to the operationalization of the main hypothesis. This can be done by using existing indicators, common definitions taken from a state of the art, or by elaborating a truth table based on a crisp-set QCA, in which all the possible cases are listed and the attributes of T and O are assessed for each possible case. The second step consists in elaborating a typology of cases for consistency, based on a fuzzy-set QCA, in which all cases are dispatched according to their class membership for the theoretical mechanism (Schneider and Rohlfing 2013). A typical case (full membership) combines both the trigger and the outcome. A deviant case (partial membership on the cause) exhibits the attributes of the trigger but not of the outcome. An individually irrelevant (no membership) case neither shows the attributes of the trigger nor of the outcome. An inconsistent case (partial membership on the effect) shows the attributes of the outcome but not of the trigger.

A typical case is best for theory-building because it presents the most likely situation where a detectable but non-observable causal mechanism actually exists, that links the observed trigger and outcome (Beach and Rohlfing 2015). Deviant cases are best for theory-testing because they allow to revise and precise the initial causal mechanism. Individually irrelevant cases are out of the scope of process tracing, since the method does not deal with symmetry in causality (that is the absence of a causal factor is not considered as a cause of the absence

Table 16.4 *Typology of cases*

		Resource nationalist oil and gas policy	
		T	-T
		Typical:	
	O	Argentina (2007–2015) Bolivia (2006–2018) Ecuador (2007–2016) Venezuela (1999–2013)	Inconsistent: Colombia (2010–2018)
Public accountability deficit			Individually irrelevant:
	-O	Deviant: Brazil (2003–2010) Mexico (2012–2018)	Peru (2001–2006), Trinidad & Tobago (1987–1997)

Source: Fontaine, Medrano, and Narváez (2020).

of an outcome). Inconsistent cases are useful inasmuch as they would reveal the inconsistency of the main hypothesis, due to spurious or omitted variables, hence leading to a new process theorization.

Table 16.4 presents a typology of cases based on the theoretical relationship linking the adoption of resource nationalism as a policy aim to a public accountability deficit in oil policy in Latin America (based on Fontaine, Medrano, and Narváez 2020).

The total population of oil and gas producing countries in Latin America during the period of reference includes nine cases among which four typical, two deviant for consistency, two individually irrelevant, and one inconsistent (which presents an omitted variable that is not observed in any other case, possibly a protracted armed conflict). The attributes of the trigger are based on a common definition of resource nationalism, which combines state-centered development with the use of oil rents for development and state control over oil rents (Haslam and Heidrich 2016). The threshold defining class membership for the trigger is based on the current legal regulation of the oil sector, with T being positive if the oil and gas sector is partially controlled by the state through national oil companies protected from the market competition by a special status. The attributes of the outcome are based on three indicators taken from the World Governance Index: voice and accountability, regulatory quality and rule of law (WGI 2018). The threshold is defined by a negative variation of these indicators during the period following the adoption of a new oil and gas policy.

After its causal homogeneity has been established thanks to the typology of cases presented above, the research can proceed with a congruence analysis of typical and deviant cases.

3.4 Step 4: Congruence Analysis

This congruence analysis is based on a comparison of at least two typical cases and two deviant cases. It is a safe way to assess the reliability of the theoretical causal mechanism and the consistency of case selection according to this mechanism. It is less demanding than a multiple-case process tracing because it does not seek to prove the necessity in a causal process; but it provides more insights than the case typology because it compares the pathways followed by multiple cases within the same analytical framework. This comparison can lead to three different conclusions: to confirm the existence of a causal mechanism, to disconfirm it, or to identify the factors of equifinality (Beach and Rohlfing 2015: 11). In our example, the

two typical cases (Venezuela and Ecuador) confirm the existence of a cross-case relationship confirming the theoretical mechanism, the two deviant cases (Brazil and Mexico) indicate that the mechanism is deviated by the existence of a transparency policy that is absent from each typical case but present in both the individually irrelevant cases and the inconsistent one (Fontaine, Medrano, and Narváez 2020).

The first typical case (Venezuela) provides conclusive evidence of a causal mechanism linking the adoption of resource nationalist policy aims to a deficit of public accountability. The second typical case (Ecuador) provides conclusive evidence confirming the existence of this mechanism, although the contingent modalities were different in degree and nature. In both cases, the government actually formulated a new oil policy based on resource nationalism. Then the incumbent reformed the cross-sectorial coordination in the sense of a major centralization and hierarchy subordinating the oil sector to the executive power. This generated protracted social conflicts that were repressed by the government in accordance with a hierarchical implementation style, which ended producing a deficit of public accountability at all levels in the oil policy.

The two deviant cases (Brazil and Mexico) provide convincing evidence of the existence of the trigger, with both governments claiming policy aims based on resource nationalism. Yet they also provide evidence of the mechanism breakout. In the case of Brazil, the government actually formulated a new oil policy based on resource nationalism but they did not manage to adapt the institutional system to this policy, essentially because of a huge scandal of corruption affecting the national oil company Petrobras (aka “Operation car wash”). In the meantime the government also implemented a policy fostering transparency and social control, which ended in improving public accountability. In the case of Mexico, the government formulated a new oil policy aimed at attracting foreign direct investments through partial privatization, rather than national development based on state intervention. They also implemented transparency and anti-corruption policies that applied to the oil sector, which ended in improving public accountability.

Once the congruence analysis has been conducted, the theoretical causal process can be partly revisited and nuanced in a deep within-case study.

3.5 Step 5: Deep Within-Case Study

To turn the expected empirical observations into evidence requires three intermediate steps: data collection, empirical tests report, and evidence assessment.

Building an index of evidence is standard operating procedure for any method but it can be challenging when the expected empirical observations include such murky and heterogeneous material as laws, statistics, technical reports, interviews, etc. A transparent way to proceed is by relating each part of the theoretical process with the collected data, their corresponding sources, and the tests results.

Table 16.5 presents a sample of the data collected on the trigger of a typical case. This table shows that a confirmation of H_T may result from four positive tests on the instruments of nodality, authority, treasure, and organization. For each test, a single primary source (as for P_{T1} , P_{T3} and P_{T4}) or a triangulation of secondary sources (as for P_{T2}) provides the information needed. For each information, the index by source allows to cross-check the existence of the information and the interpretation it is given by the author.

Table 16.5 *Evidence of T in a typical case*

Hypothesis	Expected empirical observations	Evidence found	Sources	Tests results
Government aims at securing development through resource nationalism (H_T)	Explicit claims in government's program (P_{T1})	National Plan of Good Living calling for structural reform Election of Constitutional	National Plan for 2007–2011	+
	Explicit claims by government in constitutive regulation (P_{T2})	Assembly and press reports on debates; Substitution of the Congress by a National Assembly Government to increase	Press reports on 2007–2008 constitutional process in Ecuador	+
	Explicit claims by government in development model (P_{T3})	government-take to finance development; government confiscates extraordinary utilities by private companies	Reports by legislative commission on 2008 Financial Law	+
	Explicit claims by government in state apparatus design (P_{T4})	Government creates SENPLADES	Executive decree on agency creation (2007)	+

That being said, transparency in the presentation is not sufficient to turn these data into evidence of the theoretical process, hence the usefulness of the Bayesian formalization as explained in step 2. In the policy design framework utilized in our example (Fontaine, Medrano, and Narváez 2020), the instruments of nodality, authority, treasure, and organization are treated as hoop tests, even if some of them may be doubly-decisive.⁴ The rationale is that if a policy exists there is a high probability to find traces of it in these instruments, so $p(e|h) = 0.9$, but there is still a high probability to find these fingerprints in absence of such a policy, so $p(e|\neg h) = 0.3$.

For instance, if an oil policy based on resource nationalism exists, there is a high probability to find traces of resource nationalism attributes in the mix of policy instruments. Yet finding traces of resource nationalism attributes in one instrument does not mean the oil policy is based on resource nationalism. This conservative strategy seeks to avoid the potential statistical bias due to tests sequencing (for instance performing a hoop test before or after a doubly-decisive would affect the final result of the tests series). It also spares the tedious justification of random values selection, to assess certainty and uniqueness (see Paz and Fontaine 2018).

The replication of independent hoop tests on the four types of instruments used in the policy design framework from our example allows to upgrade or downgrade the main hypothesis for each part of the causal process. Table 16.6 presents a simulation of such tests results based on $p(h) = 0.5$.

After four hoop tests, based on a 0.5 prior, the posterior confidence varies from 0.99 (with 4/4 positive tests) to 0.79 (with 3/4 positive tests), 0.16 (with 2/4 positive tests), and 0.01 (with 1/4 positive test). This means if only 1/4 test was negative or inconclusive the corresponding part of the main hypothesis – hence the overall theoretical process – would remain indeterminate, and if at least 2/4 tests were negative or inconclusive, the prior confidence should be downgraded.

Table 16.6 Probative value of hoop tests on policy instruments

Empirical tests	Tests result	p(h)	p (¬h)	p (e h)	p (¬e h)	p (e ¬h)	p (¬e ¬h)	p (h e)	p (h ¬e)
P _t 1	+	0.50	0.50	0.90	0.10	0.30	0.70	0.75	0.13
P _t 2	+	0.75	0.25	0.90	0.10	0.30	0.70	0.90	0.30
P _t 3	+	0.90	0.10	0.90	0.10	0.30	0.70	0.96	0.56
P _t 4	+	0.96	0.04	0.90	0.10	0.30	0.70	0.99	0.79
P _A 1	-	0.50	0.50	0.90	0.10	0.30	0.70	0.75	0.13
P _A 2	+	0.13	0.88	0.90	0.10	0.30	0.70	0.30	0.02
P _A 3	+	0.30	0.70	0.90	0.10	0.30	0.70	0.56	0.06
P _A 4	+	0.56	0.44	0.90	0.10	0.30	0.70	0.79	0.16
P _B 1	-	0.50	0.50	0.90	0.10	0.30	0.70	0.75	0.13
P _B 2	-	0.13	0.88	0.90	0.10	0.30	0.70	0.30	0.02
P _B 3	+	0.02	0.98	0.90	0.10	0.30	0.70	0.06	0.00
P _B 4	+	0.06	0.94	0.90	0.10	0.30	0.70	0.16	0.01
P _C 1	-	0.50	0.50	0.90	0.10	0.30	0.70	0.75	0.13
P _C 2	-	0.13	0.88	0.90	0.10	0.30	0.70	0.30	0.02
P _C 3	-	0.02	0.98	0.90	0.10	0.30	0.70	0.06	0.00
P _C 4	+	0.00	1.00	0.90	0.10	0.30	0.70	0.01	0.00

Source: Fontaine, Medrano, and Narváez (2020).

Once the expected empirical observations have been converted into confirming or disconfirming evidence for the theoretical causal process, the research can proceed to another deep within-case study, aimed at comparing the first case with another typical case or with a deviant case for theory-testing, based on the typology built in step 3.

The five-step protocol presented in this section shows how realist process tracing rolls out from causal process theorization to cross-case comparison through the design of precise empirical tests, transparent case selection, and formal assessment of collected evidence.

4. CONCLUSION: PROCESS TRACING FOR POLICY LEARNING

This chapter has argued against a neo-positivist standpoint on process tracing, in order to value its contribution to comparative policy analysis. A distinction between philosophical and scientific ontologies was first established to make a difference between methodology and methods or techniques. Based on this distinction, realism has been defined as a methodology combining dualism with transfactualism. Then a definition of realist process tracing as a method has been proposed, which places social mechanisms at the heart of a causal process, as non-observable but detectable forces.

In the second part, the chapter proceeded with describing the standard operating procedures to apply realist process tracing to comparative policy analysis, based on research about the causal relationship between resource nationalism and public accountability deficits in Latin America. First, policy design served as a theoretical framework to set up a causal process leading from the adoption of a policy aim to the production of a policy outcome, through a threefold mechanism including policy formulation, inter-sectorial coordination, and political interplays. Second, a typology of cases for consistency was elaborated, to define typical and deviant cases for selection and comparison. Third, a series of 20 hoop tests were designed,

using policy instruments as expected empirical observations. Fourth, a standard for information gathering and tests results was set, before explaining how to use the Bayes theorem of likelihood to turn these empirical observations into evidence. Finally the chapter explained how to conduct a cross-case comparison to assess the theoretical causal process, eventually leading to reformulate the causal mechanism after addressing contingency.

Realist process tracing is time-consuming but this method offers a way to replicate the analysis of a causal process without sacrificing transparency and precision. It should also constitute an incentive to improve policy design through social learning.⁵

NOTES

1. As of mid-2018, this book registered 9,491 citations on Google Scholars, which is to be compared to classical textbooks on case study methods such as *Designing Social Inquiry* by King, Keohane, and Verba (1994) (10,042 records), *Case Study Research* by Gerring (2007a) (5,326 records) and *Rethinking Social Inquiry* by Brady and Collier (2010) (1,625 records).
2. We refer to T and O instead of X and Y, in order to avoid any confusion with the probabilistic language of independent, dependent, and intervening variables used in neo-positivist process tracing.
3. NB: The values of $p(e/h)$ and $p(e/\neg h)$ are consistent with the scale of confirming/disconfirming values of each kind of test. A positive result of a straw-in-the-wind test cannot increase the prior confidence in a way that is superior to a positive result of the other kinds of tests; and vice versa, a positive doubly-decisive test cannot increase the prior confidence in a way that is inferior to the others. In any case, a confirming test implies that $p(h/e) > p(h)$, so the values cannot produce a negative result if the test is positive. Finally $p(e/\neg h) > 0.5$ would mean e is inconsistent with h, so this value should be ≤ 0.5 for the expected empirical observation to be relevant.
4. For instance treasure instruments could provide doubly-decisive tests, since there is a high probability to find a treasure instrument in any policy, while finding a treasure instrument in absence of a policy would be unexpected.
5. This chapter owes a lot to the works by the research group on comparative public policy at FLACSO Ecuador. I am particularly grateful to Iván Narváez for his constant support and collaboration. My gratitude goes to Derek Beach, for being the mentor of the debate, to Patrick T. Jackson for initiating me to realism, and to B. Guy Peters, Jale Tosun and Keith Dowding for their comments on the first draft. All eventual errors are my own responsibility.

REFERENCES

- Archer, M., Bhaskar, R., Collier, A., Lawson, T., and Norrie, A. (Eds.) (1998). *Critical Realism: Essential Readings*. London: Routledge.
- Astbury, B. and Leeuw, F. L. (2010). Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation* 31 (3): 363–81.
- Beach, D. (2013). Taking mechanisms seriously? *European Political Science* 12 (1): 13–15.
- Beach, D. (2016). It's all about mechanisms: what process tracing case studies should be tracing. *New Political Economy* 21 (5): 463–72.
- Beach, D. (2017). Achieving methodological alignment when combining QCA and process tracing in practice. *Sociological Methods & Research* 47 (1): 64–99.
- Beach, D. (2018). Multi-method research in the social sciences: a review of recent frameworks and a way forward. *Government and Opposition*. doi:10.1017/gov.2018.53.
- Beach, D. and Pedersen, R. B. (2013). *Process Tracing Methods: Foundations and Guidelines*. Ann Arbor, MI: University of Michigan Press.
- Beach, D. and Pedersen, R. B. (2016a). *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching and Tracing*. Ann Arbor, MI: University of Michigan Press.

- Beach, D. and Pedersen, R. B. (2016b). Selecting appropriate cases when tracing causal mechanisms. *Sociological Methods & Research* 47 (4): 837–71.
- Beach, D. and Rohlfing, I. (2015). Integrating cross-case analyses and process tracing in set-theoretic research: strategies and parameters of debate. *Sociological Methods & Research* 47 (1): 3–36.
- Befani, B. and Mayne, J. (2014). Process tracing and contribution analysis: a combined approach to generative causal inference for impact evaluation. *IDS Bulletin* 45 (6): 17–36.
- Befani, B. and Stedman-Bryce, G. (2016). Process tracing and Bayesian updating for impact evaluation. *Evaluation* 23 (1): 42–60.
- Bennett, A. (2006). Stirring the frequentist pot with a dash of Bayes. *Political Analysis* 14 (3): 339–44.
- Bennett, A. (2015). Appendix: Disciplining our conjectures: systematizing process tracing with Bayesian analysis. In A. Bennett and J. Checkel (Eds.), *Process Tracing: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press, 276–98.
- Bennett, A. and Checkel, J. (Eds.) (2015a). *Process Tracing: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press
- Bennett, A. and Checkel, A. (2015b). Process tracing: from philosophical roots to best practices. In A. Bennett and J. Checkel (Eds.), *Process Tracing: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press, 3–38.
- Bennett, A. and Elman, C. (2006). Complex causal relations and case study methods: the example of path dependence. *Political Analysis* 14 (3): 250–267.
- Bhaskar, R. (2008 [1978]). *A Realist Theory of Science*. London: Routledge.
- Blatter, J. and Haverland, M. (2012). *Designing Case Studies: Explanatory Approaches in Small-N Research*. Basingstoke: Palgrave Macmillan.
- Blatter, J. and Haverland, M. (2014). Case studies and (causal-) process tracing. In I. Engeli and C. Rothmayr Allison (Eds.), *Comparative Policy Studies: Conceptual and Methodological Challenges*. Basingstoke: Palgrave Macmillan, 59–84.
- Brady, H. and Collier, D. (Eds.) (2010). *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield.
- Brady, H., Collier, D., and Seawright, J. (2010). Refocusing the discussion of methodology. In H. Brady and D. Collier (Eds.), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield, 15–32.
- Charbonneau, E., Henderson, A. C., Ladouceur, B., and Pichet, P. (2016). Process tracing in public administration: the implications of practitioner insights for methods of inquiry. *International Journal of Public Administration* 40 (5): 434–42.
- Collier, D. (2011). Understanding process tracing. *Political Science and Politics* 44 (4): 823–30.
- Crasnow, S. (2017). Process tracing in political science: what’s the story? *Studies in History and Philosophy of Science* 62: 6–13.
- Fairfield, T. and Charman, A. (2017). Explicit Bayesian analysis for process tracing: guidelines, opportunities, and caveats. *Political Analysis* 25 (3): 363–80.
- Falleti, T. (2016). Process tracing of extensive and intensive processes. *New Political Economy* 21 (5): 455–62.
- Falleti, T. and Lynch, J. (2009). Context and causal mechanisms in political analysis. *Comparative Political Studies* 42 (9): 1143–66.
- Fontaine, G., Fuentes, J. L., and Narváez, I. (2018). Policy mixes against oil dependence: resource nationalism, layering and contradictions in Ecuador’s energy transition. *Energy Research and Social Science* 47: 56–68.
- Fontaine, G., Medrano, C., and Narváez, I. (2020). *The Politics of Public Accountability: Policy Design in Latin American Oil Exporting Countries*. Cham: Palgrave Macmillan.
- Fontaine, G., Narváez, I., and Paz, B. (2017). Explaining public accountability deficit in extractive policies: the Ecuadorian case. *Extractive Industries and Society* 5 (1): 190–200.
- Geddes, B. (2003). *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics*. Ann Arbor, MI: University of Michigan Press.
- George, A. and Bennett, A. (2005). *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- George, A. L. and McKeown, T. J. (1985). Case studies and theories of organizational decision making. *Advances in Information Processing in Organizations: A Research Annual* 2: 21–58.

- Gerring, J. (2007a). *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.
- Gerring, J. (2007b). The mechanistic worldview: thinking inside the box. *British Journal of Political Science* 38 (1): 161–79.
- Goertz, G. and Mahoney, J. (2012). *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.
- Goertz, G. and Starr, H. (2003). Introduction: necessary condition logics, research design, and theory. In G. Goertz and H. Starr (Eds.), *Necessary Conditions: Theory, Methodology, and Applications*. Lanham, MD: Rowman & Littlefield, 1–24.
- Hall, P. (1993). Policy paradigms, social learning, and the state: the case of economic policymaking in Britain. *Comparative Politics* 25 (3): 275–96.
- Hall, P. (2003). Aligning ontology and methodology in comparative politics. In J. Mahoney and D. Rueschemeyer (Eds.), *Comparative Historical Analysis in the Social Sciences*. New York: Cambridge University Press, 373–404.
- Hall, P. (2013). Tracing the progress of process tracing. *European Political Science* 12 (1): 20–30.
- Haslam, P. and Heidrich, P. (2016). From neoliberalism to resource nationalism: states, firms and development. In P. Haslam and P. Heidrich (Eds.), *The Political Economy of Natural Resources and Development: From Neoliberalism to Resource Nationalism*. London: Routledge, 1–32.
- Hedström, P. and Swedberg, R. (Eds.) (1998). *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
- Hedström, P. and Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology* 36: 49–67.
- Hempel, C. and Oppenheim, C. (1948). Studies in the logic of explanation. *Philosophy of Science* 15 (2): 135–75.
- Hood, C. (1986). *The Tools of Government*. London: Macmillan.
- Hood, C. (2007). Intellectual obsolescence and intellectual makeovers: reflections on the tools of government after two decades. *Governance* 20 (1): 127–44.
- Hood, C. and Margetts, H. (2007). *The Tools of Government in the Digital Age*. New York: Palgrave Macmillan.
- Howlett, M. (2009). Governance modes, policy regimes and operational plans: a multi-level nested model of policy instrument choice and policy design. *Political Sciences* 42: 73–89.
- Howlett, M. (2011). *Designing Public Policies: Principles and Instruments*. New York: Routledge.
- Howlett, M. and Rayner, J. (2006). Understanding the historical turn in the policy sciences: a critique of stochastic, narrative, path dependency and process-sequencing models of policy-making over time. *Policy Sciences* 39: 1–18.
- Humphreys, M. and Jacobs, A. (2015). Mixing methods: a Bayesian approach. *American Political Science Review* 109 (4): 653–73.
- Jackson, P. T. (2016). *The Conduct of Inquiry in International Relations: Philosophy of Science and its Implications for the Study of World Politics* (2nd edition). London: Routledge.
- Kay, A. (2005). A critique of the use of path dependency in policy studies. *Public Administration* 83: 553–71.
- Kay, A. and Baker, P. (2015). What can causal process tracing offer to policy studies? A review of the literature. *Policy Studies Journal* 43: 1–21.
- Kern, F., Kuzemko, C., and Mitchell, C. (2014). Measuring and explaining policy paradigm change: the case of UK energy policy. *Policy & Politics* 42 (4): 513–30.
- King, G., Keohane, R., and Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Linder, H. S. and Peters, B. G. (1998). The study of policy instruments: four schools of thought. In B. G. Peters and F. K. M. van Nispen (Eds.), *Public Policy Instruments: Evaluating the Tools of Public Administration*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing, 33–45.
- Lorentzen, P., Fravel, M. T., and Paine, J. (2016). Qualitative investigation of theoretical models: the value of process tracing. *Journal of Theoretical Politics* 29 (3): 467–91.
- Machamer, P. (2004). Activities and causation: the metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science* 18 (1): 27–39.
- Mackie, J. (1965). Causes and conditions. *American Philosophical Quarterly* 2 (4): 245–64.

- Mahoney, J. (2001). Beyond correlational analysis: recent innovations in theory and method. *Sociological Forum* 16 (3): 575–93.
- Mahoney, J. (2016). Mechanisms, Bayesianism, and process tracing. *New Political Economy* 21 (5): 493–9.
- Mahoney, J. and Goertz, G. (2006). A tale of two cultures: contrasting quantitative and qualitative research. *Political Analysis* 14 (3): 227–49.
- Marsh, D. and Furlong, P. (2010). A skin, not a sweater: ontology and epistemology in political science. In G. Stoker and D. Marsh (Eds.), *Theory and Methods in Political Science*. Basingstoke: Palgrave Macmillan, 184–211.
- McKay Illari, P. (2011). Mechanistic evidence: disambiguating the Russo–Williamson thesis. *International Studies in the Philosophy of Science* 25 (2): 139–57.
- Morgan, K. (2016). Process tracing and the causal identification revolution. *New Political Economy* 21 (5): 489–92.
- Paquet, M. and Broschek, J. (2017). This is not a turn: Canadian political science and social mechanisms. *Canadian Journal of Political Science / Revue canadienne de science politique* 50 (1): 295–310.
- Paz, B. and Fontaine, G. (2018). A causal mechanism of policy innovation: the reform of Colombia’s oil-rents management system. *Revista de Estudios Sociales* 63: 2–19.
- Peters, B. G. (2013). *Strategies for Comparative Research in Political Science*. Basingstoke: Palgrave Macmillan.
- Peters, B. G. (2018). *Policy Problems and Policy Design*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Popper, K. (2014 [1959]). *The Logic of Scientific Discovery*. London: Routledge.
- Ringeling, A. B. (2005). Instruments in four: the elements of policy design. In P. Eliadis, M. Hill, and M. Howlett (Eds.), *Designing Government: From Instruments to Governance*. Montreal: McGill–Queen’s University Press, 185–202.
- Rohlfing, I. (2013). Varieties of process tracing and ways to answer why-questions. *European Political Science* 13: 31–9.
- Rohlfing, I. (2014). Comparative hypothesis testing via process tracing. *Sociological Methods & Research* 43 (4): 606–42.
- Rohlfing, I. and Schneider, C. (2016). A unifying framework for causal analysis in set-theoretic multi-method research. *Sociological Methods & Research* 47 (1): 37–63.
- Ruback, T. J. (2010). “Let me tell the story straight on”: Middlemarch, process-tracing methods and the politics of narrative. *British Journal of Politics and International Relations* 12: 477–97.
- Russo, F. and Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science* 21 (2): 157–70.
- Sayer, A. (2000). *Realism in Social Science*. London: Sage.
- Sayer, A. (2010). *Method in Social Science: A Realist Approach*. London: Routledge.
- Schneider, C. and Rohlfing, I. (2013). Combining QCA and process tracing in set-theoretic multi-method research. *Sociological Methods & Research* 42 (4): 559–97.
- Seawright, J. (2016). *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*. Cambridge: Cambridge University Press.
- Seawright, J. and Gerring, J. (2008). Case selection techniques in case study research. *Political Research Quarterly* 61 (2): 294–308.
- Skocpol, T. and Somers, M. (1980). The uses of comparative history in macrosocial inquiry. *Comparative Studies in Society and History* 22 (2): 175–97.
- Snow, J. (1855). *On the Mode of Communication of Cholera*. London: John Churchill.
- Stinchcombe, A. (1991). The conditions of fruitfulness of theorizing about mechanisms in social science. *Philosophy of the Social Sciences* 21 (3): 367–88.
- Trampusch, C. and Palier, B. (2016). Between X and Y: how process tracing contributes to opening the black box of causality. *New Political Economy* 21 (5): 437–54.
- Van Evera, S. (1997). *Guide to Methods for Students of Political Science*. New York: Cornell University Press.
- Wendt, A. (1999). *Social Theory of International Politics*. Cambridge: Cambridge University Press.
- WGI (Worldwide Governance Indicators 2018). Open access database available at: <http://info.worldbank.org/governance/wgi/#home>.

PART V

QUALITATIVE TECHNIQUES

17. Using focus groups in comparative policy analysis

Patrik Marier, Daniel Dickson and Anne-Sophie Dubé

1. INTRODUCTION

Focus group methods have been in use over 100 years, but grew in popularity after World War II when they were in vogue in communication research and propaganda analysis (Merton, Fiske, and Kendall 1990 [1956]). In social research, focus groups experienced a revival in the 1990s (Wilkinson 1998b). Today, they are utilized across a wide range of disciplines, and are particularly prominent in marketing studies with over a quarter of a million focus groups deployed every year in the US in this field of study (Stewart and Shamdasani 2015). However, the use of focus groups in public policy scholarship remains rare relative to other research methods such as survey research, individual interviews, and statistical analyses, and even more so in comparative policy analysis. A search for comparative public policy articles in three leading journals (*Governance*, *Journal of Comparative Policy Analysis*, and *Journal of European Public Policy*) reveals only four articles where the authors conducted focus groups (Dimitrova and Kortenska 2017; Gleeson et al. 2011; Koenig-Archibugi and Macdonald 2013; Stapenhurst and Pelizzo 2012)¹ and only one of them compares across countries (Dimitrova and Kortenska 2017). As we explain later in this chapter, the cost of conducting focus groups, in terms of both time and money, is a key reason why this method is not more popular in comparative policy analysis.

The aim of this chapter is to provide a succinct overview of the use of focus group methods in comparative policy research, and in other social sciences. In addition, it provides insights into the methodological benefits and challenges of using focus groups for comparative policy research. For this portion, the chapter relies strongly on the experience of eight focus groups conducted in the province of Québec (Canada), as part of a large comparative project analyzing the generosity of social policies targeting older adults. There is already a rich literature on how to conduct focus groups (see for example, Stewart and Shamdasani 2015), we focus instead on why these methods should be used in (comparative) public policy and on offering lessons from the recent experience of our research team.

This chapter is structured as follows. In section 2, we provide an operational definition of focus group methods, and discuss the reasons why comparative public policy scholars would want to employ them. In section 3, we outline the rationale as to why we chose to deploy focus groups in Québec for the project on social policies for older adults. The main methodological advantages of focus groups follow in section 4, which we embolden with practical examples from our focus group experiences. Conversely, section 5 discusses some of the challenges with focus group methods, based both on the literature and our experiences in the field. Sections 6 and 7 discuss the issues of designing focus groups and managing focus group dynamics respectively. In sum, we aim to provide a clearer picture of when focus group methods are

appropriate in comparative policy analysis so as to encourage their increasing use, particularly in exploratory research.

2. WHY USE FOCUS GROUPS IN COMPARATIVE PUBLIC POLICY?

Prior to discussing focus groups, it is imperative to define what a focus group method entails. According to Morgan (1996), a focus group involves three important components: (1) its primary objective is to collect data; (2) it puts the emphasis on the interaction amongst members of the focus groups; and (3) it presupposes an active role on the part of the researcher in the creation of the focus group (Morgan 1996: 130; see also Wilkinson 1998b). Most commonly, focus groups involve the assemblage of six to ten participants in one room with discussion topics led by a moderator; however, alternative methods such as telephone and online focus groups involving wide ranges of participants are common practice in marketing research (Litosseliti 2003: 6). Social scientists have most commonly used focus group methods in the exploratory stages of research because they provide access to a greater number of participants than individual interviews or ethnographic research, while still allowing participants to form responses using their own words, with less influence from categories or associations imparted by the researcher (Stewart, Shamdasani, and Rook 2007: 40). This is particularly beneficial when dealing with contested or multi-dimensional concepts, which may require a deeper descriptive foundation in exploratory phases to enable hypothesis building. However, as we will demonstrate in the discussion of advantages and disadvantages of focus groups, realizing these benefits from focus group methods requires significant prior organization by the research team, and discipline by the moderator when facilitating.

There are at least three key reasons as to why scholars should consider the use of focus groups in their research design in comparative public policy. The first reason is simply to provide a better understanding of how individuals interpret the policy environment they live in. An underlying assumption of the policy literature is, simply put, that policy matters (Hecl 1972). As such, it assumes, for example, that policies and policy instruments target segments of the populations differently (Schneider and Ingram 1993). To this end, focus groups have been extremely valuable to provide more in-depth analyses of the impact of policy tools on specific groups such as women on welfare (Hagen 1994) and people who inject drugs (Lancaster et al. 2015). In this way, focus group methods can be beneficial to researchers who seek to compare the outcomes of specific policies across different groups and to determine where benefits and burdens occur within a policy context.

Current policies also impact future policy development since they shape the conceptualization of policy problems and, consequently, the politics of said policy or program (Béland 2005; Pierson 1993). While focus groups are unlikely to yield substantive insight into how policies are developed, since individual interviews with key policy players are more likely to produce better results, they can be extremely useful to analyze how groups of individuals interpret or respond to policies. For example, in Gleeson et al. (2011), a focus group discussion with eight policy practitioners in the department of human services in Australia helped to identify strategies to build organizational capacity quickly in the wake of an administrative reform. The deployment of a focus group, following a first stage of research featuring individual inter-

views, led to the identification of complementary issues that provided a richer understanding of ongoing tensions surrounding capacity building.

The second principal reason is to enrich the quality of the information and the data gathered. As such, focus groups are often used to complement other methods and are rarely used as a stand-alone method (Wilkinson 1998b). Indeed, this multi-method approach to focus group usage is common in social sciences. A review article reports that 60% of all articles using focus groups from the Sociological Abstracts employed another research method as well (Morgan 1996: 130). The use of focus groups, as a complementary method, typically occurs in the initial phase of a research project to generate hypotheses and at the very end to add depth to the research materials, most notably to provide a more detailed understanding of the research subject under scrutiny (Wilkinson 1998b: 184–5). This is not to suggest that focus groups can never be used as a singular data source, but rather that they are most advantageous to researchers who require thicker description, particularly in concert with alternative methods. To this end, researchers have embraced the potential of focus groups in triangulation research designs, specifically when used alongside individual interviews to draw upon to relative advantages of each method (Caillaud and Flick 2017).

From the four comparative public policy articles cited in the introduction, all combined individual interviews with focus groups (Dimitrova and Kortenska 2017; Gleeson et al. 2011; Koenig-Archibugi and Macdonald 2013; Stapenhurst and Pelizzo 2012). For example, in their analysis on the benefit of parliamentary oversight in Ghana, Stapenhurst and Pelizzo (2012) conducted 52 face-to-face interviews and utilized four focus groups for each type of policy actors involved in the research project (politicians, parliamentary staff, civil society representative, and journalist). Interestingly, they concluded their research with a fifth focus group from a mix of policy actors amongst the 52 interviewed earlier. The purpose of this focus group was to validate the preliminary analysis and the conclusions of the research (Stapenhurst and Pelizzo 2012: 336). In so doing, they draw upon the advantages of focus groups in analyzing and comparing interaction effects, both in homogeneous and heterogeneous group contexts.

This leads to a third reason to use focus groups in comparative public policy, which, in the spirit of comparison, is to assess views and opinions of diverse segments of the population (Dimitrova and Kortenska 2017) or policy actors (Stapenhurst and Pelizzo 2012). Comparative cases in focus group studies are often set within single jurisdictions or regions, and comparison across countries is quite rare, as this type of study “adds substantially to the already considerable time, effort, and funds” required (Knodel 1995: 11). Nonetheless, in a comparative analysis of discourse concerning European Union enlargement, Dimitrova and Kortenska (2017) deployed focus groups as part of the Q methodology (see McKeown and Thomas 2013) to contrast the discourses on enlargement in two old and recent member states. In contrast to many contributions within discursive institutionalism, the primary focus is on citizen discourse (Dimitrova and Kortenska 2017: 261). The researchers extracted discourses on enlargement from three to five focus groups for each of the four countries (Germany, the Netherlands, Poland, and Bulgaria). This resulted in the creation of 64 statements on enlargement that were later utilized for individualized interviews (40 per country). The responses were coded and investigated further via factor analysis (Dimitrova and Kortenska 2017: 264). In this application focus groups were used to expediently identify country-specific perspectives, reflecting a unique group consciousness that could be compared with respect to an overarching policy issue. We adopted a similar approach, albeit on a smaller scale, to the use of focus groups in our study.

3. THE DEPLOYMENT OF FOCUS GROUPS IN QUÉBEC

This section summarizes the motivation for the use of focus groups as part of a five-year research project analyzing the generosity of social policies, including social services, for older adults across 20 industrialized countries.² The project employs a variety of research methods (including statistical analysis, survey, and individual interviews). The project involves three distinct phases: (1) focus groups with older adults; (2) the construction of a social policy index; and (3) field work including individual interviews with policy makers in five countries (Canada, France, USA, South Korea, and Sweden).

The inspiration to deploy focus groups originates from the traditional practices utilized by the Centre for Research and Expertise in Social Gerontology (CREGÉS) where one of the authors is scientific director since 2013 (Marier). The research center aims to encourage research endeavors that are embedded in practice, working closely with health and social services providers. CREGÉS also has a long-standing tradition of conducting participatory action research with older adults. As such, older adults participate actively in the research and knowledge transfer activities of CREGÉS. Focus groups remain a popular methodological tool with projects involving participatory action research to access the views of groups that are traditionally marginalized (Wilkinson 1998b: 185). This is the case of older adults, for example, who experience various forms of ageism in their relationship with professionals when seeking to access various services to sustain a good quality of life as they age (Chrisler, Barney, and Palatino 2016; Koch and Webb 1996; São José et al. 2019). Older adults also have a different understanding than professionals and service providers as to what constitutes successful aging (Bowling 2007) or even autonomy, a core concept in the assessment of needs for older adults (Welford et al. 2012). As a result of benefiting from a mandate to develop innovative approaches through research, combined with a tradition to involve older adults and practitioners, CREGÉS researchers have been using focus groups frequently in their research projects. This includes, recently, contributions on how to encourage residents and patients to engage in care planning in long-term care (Sussman et al. 2017), and the development of a new screening tool for older adult mistreatment (Couture et al. 2019).

Health research dominates the field of aging. Social gerontologists have been amongst the most notable critics of the “biomedicalization of aging” and its multiple consequences (see, for example, Estes and Binney 1989; Kaufman 2004; Koch and Webb 1996). Hence, to capture the social realities of aging, the research team opted to operate in a forum where there would be an opportunity to encourage discussions amongst older adults on what a *social* approach to aging entails. Early findings clearly support this decision since references to health were ubiquitous despite the fact that our questionnaire involved only social policy elements (Marier, Dickson, and Dubé 2017). It was through participant engagements that the social elements were brought to light, which even led in some cases to an awakening to the importance of prioritizing the social dimensions of aging. Alternative methods of inquiry to pool older adults, such as survey, would not have achieved similar results.

Another motivating element justifying the use of a focus group in this current research project is simply that the use of this methodological tool is far more common and accepted in social gerontology (Knodel 1995) than in public policy. A quick survey of the literature reveals recent articles using focus groups on a wide range of social issues affecting older adults such as driving (Adler and Rottunda 2006; Yassuda, Wilson, and Mering 1997), aging in place (Emlet and Mocerri 2012; Walsh and O’Shea 2008; Wiles et al. 2012), the intersection between

aging and gender (Krekula 2007), sexuality and aging (Orel 2004), quality of life (Higgs et al. 2003; Hyde et al. 2003), and work retention (Claes and Heymans 2008). Hence, the use of focus groups opens opportunities to engage with other comparable research findings.

It is noteworthy that all of the issues addressed by the social gerontological literature here reviewed pertain to policy areas for older adults. Moreover, these articles demonstrate that focus group research is well-suited to complement other research strategies in identifying best practices in policy interventions for this specific population. For example, Wiles et al. (2012) employ focus groups alongside individual interviews to find that home care policies promoting aging in place tend to define “place” too narrowly as a specific domicile, rather than the broader community with which older adults are familiar. This has important implications for home care policy, which tends to focus more on the provision of care within the domicile than in the broader community.

Similarly, Orel’s (2004) focus group and in-depth interview research with gay, lesbian, and bisexual older adults identified key implications for this sub-group, which informed the development of a new needs assessment instrument taking into account their unique concerns. As with our project, her focus group research demonstrated a surprising familiarity with the language and substance of existing assessment instruments, which are the key determinants for the receipt of social services. This lends credence to the notion, put forth by Wutich et al. (2010), that focus groups are more effective than individual response formats in stimulating responses concerning sensitive policy topics when respondents see the opportunity to contribute constructively to knowledge sharing about an important issue. In this way, group discussions provide a context that empowers individuals to respond in greater depth about how to improve existing policy mechanisms, such as assessment instruments.

4. METHODOLOGICAL ADVANTAGES

Focus groups provide the advantage of greater sample size by comparison to one-on-one interviews. That is, by engaging with a group of individuals, researchers allow for greater variation in terms of responses, thus increasing the breadth of observation for a given discussion topic. However, this advantage is offset by the decreased amount of time afforded to each individual participant, which reduces the potential for deep individual engagement on discussion topics. Focus groups should not be significantly longer in duration than one-on-one interviews, meaning that the average amount of data generated per participant will be significantly smaller than the sum of individual interviews with each participant. While this can have the effect of decreasing the amount of questions that are discussed, it does not necessarily imply less depth of discussion. Rather, focus group methods can increase descriptive depth by allowing participants to interact in an iterative process that can have the effect of empowering them and encouraging reflection and critical thinking (Litosseliti 2003). This interaction is the most important distinguishing aspect of focus group methods, and it underlies many of the advantages that these methods provide.

To understand the pivotal role that participant interaction plays in focus group research, it is useful to think of interaction as a product, rather than a process, generated by participants. The interaction that is produced in focus groups flows from the ability of participants to explain their perspectives to each other, and to request elaboration from other participants, which creates a data set that reveals important insights that are unattainable by aggregating separate

responses (Morgan 1996: 139). How individuals interact with each other in the group setting is shaped by a complex set of factors, which creates a context that is difficult to divorce from the data that is produced. This has been called the “group effect” in focus group research (Carey 1994), where the characteristic dynamics of interaction between individuals significantly influences the data. Even among the most experienced researchers, the group effect creates methodological disadvantages such as group-think or censoring effects. However, before exploring these methodological disadvantages, it is important to discuss the unique advantages that the group effect has in elevating focus group methods over other qualitative research methods such as the individual interview. For comparative policy researchers, the group effect is conducive to demonstrating the existence of, and rationale for, underlying support of or disapproval towards policies among affected social groups. In this way, observing the interactions of group members who identify with the same policy public can reveal unique aspects of group consciousness, which may enable fruitful comparative research, across policies and across publics.

4.1 Revealing Group Consciousness

When studying specific social groups, specifically those which are disadvantaged or marginalized, it is helpful for researchers to gain insight into common experiences, attitudes, or beliefs among group members. To this end, focus groups are effective in uncovering group consciousness by exploring the resonance of certain important concepts, and demonstrating the degree of consensus with respect to group beliefs within a target population, particularly in the exploratory stages of research (Rogers 2013: 232–3). Indeed, in these early stages of research, social scientists aim to assess the appropriateness of key concepts within key cultural or social contexts under study. Gerring (2001: 40) calls this the “resonance” of a concept within specialized contexts, which he lists among eight key criteria of conceptual goodness. Group interaction, the key product of focus group methods, is highly advantageous to assessing the resonance of concepts within social groups because it can give researchers insight into social contextual factors that are uniquely experienced and shared by group members.

However, it is important to emphasize that these insights do not correspond to any singular underlying objective truth, particularly with respect to the isolated responses of individual participants who may be significantly influenced by other participants. For this reason, Hollander (2004: 631) argues that focus groups are better conceptualized as research sites, rather than instruments, because they perform best as a venue to observe social interaction, rather than a means to observe individual positions. As such, focus groups give researchers unique insight into how concepts are uniquely understood by specific social groups by allowing them to observe precise processes of meaning construction. With respect to policy analysis, this can enable comparison between how different social groups similarly or differently understand the vital concepts of a specific policy problem.

We designed our focus groups to test the resonance of several concepts related to social services use among older adults in the province of Québec, including their autonomy as service users, and the generosity and coverage of existing services. To this end, we conducted eight focus groups across the province, with inter-group variation in terms of language (two conducted in English, six in French) and intra-group variations in terms of characteristics of urban and rural residency, ethnic background, and physical impairment. The purpose of the focus groups was to elucidate key concepts that would be included in a subsequent survey

instrument distributed to a broad sample of older adults in Québec. As such, the focus groups were exploratory, and were not undertaken to provide generalizable results. Participants were recruited from community groups for older adults in both Greater Montreal and Québec City. Prospective community groups were selected by contact with the largest provincial organization for social clubs across the province and by contacts associated with the research center (CREGÉS).

Once clubs were selected, we worked in conjunction with club leadership to recruit mixed-gender focus groups, ranging in size from 6–12 participants. The focus groups were conducted on site during regular club hours in order to make participation as convenient and comfortable as possible. Each group was moderated by one member of the research team, with another member present during the session to oversee audio recording and take notes. The observations of the second researcher provided valuable insights into the nature of the shifts between discussion topics, as well as the general level of group engagement that would have been difficult to observe for the moderator alone.

When analyzing the focus groups, we were able to observe subtle differences in the construction of group consciousness among differently composed groups of service users, which we later used to strengthen our survey questions. First, some of the topics that emerged consistently in group discussions appeared more important to understanding the preferences of older adult service users within the province. For example, there was a far greater emphasis on transportation services than we expected, particularly among urban participants, who tended to assign these services a higher priority than home care or respite services. Similarly, in the English-speaking groups there was discussion about how French-only customer service provision can make social services less accessible due to language barriers in information access. Both of these emergent themes demonstrated important internal variation among service users that we later attempted to tap by our survey.

In addition, our focus groups provided us with thicker description of expected phenomena, such as the paradox between preferring informal services provided by family members, and the expectation of greater formal service delivery to relieve the family care-giving burden. By deepening our understanding of this paradox in the exploratory phase of our project, we were able to design survey questions that more sensitively addressed it in the second phase of the project. As such, the mapping of a contested concept within our eight focus groups contributed to the design of key questions in our province-wide survey, and thus allowed us to test whether this care provision paradox is a more general phenomenon. This demonstrates a key advantage of the focus group method, where as researchers we were able to observe the interactional processes of constructing group consciousness regarding who should be responsible for care provision. As a result, rather than being left with two competing and incompatible narratives – government should provide versus family should provide – we were instead able to develop specific insight into when and why one narrative generates more consensus over the other. The potential to generate these kinds of unique insights in exploratory research thus stands out as a strong advantage of focus group methods.

4.2 Epistemological Implications

The nature of focus groups makes them more attractive to some epistemological perspectives than others within comparative policy analysis. Stewart, Shamdasani, and Rook (2007: 112) identify three epistemological perspectives from which focus group research proves to be

especially valuable: (i) social constructivism, (ii) phenomenology, and (iii) interpretivist approaches that accept the positions of the first two, while refusing to uncritically accept responses of participants at face value. Interpretivists, drawing from ethnography, prefer instead to look for meaning in more ephemeral clues such as body language and other expressive cues. What is consistent to all these approaches is a commitment to the importance of social context to understanding individual attitudes and behaviors. However, as Wilkinson (1998a) demonstrates, this social context is very rarely aggregated from individual responses when focus group researchers report their findings. She finds that, despite the fact that the most frequently cited advantage of the focus group methodology is in capturing group interaction, these interactions are very seldom reported in research findings. For constructivist researchers, these social interactions are invaluable in demonstrating exact processes of how meaning is socially constructed, and should thus be observed and reported as such (Wilkinson 1998a: 122). In this way, constructivist researchers are well-served by analyzing power relations within their focus groups, to elucidate mechanisms at play in meaning construction. However, the value of this approach to focus groups is predicated on the epistemological foundations, such that differently oriented researchers will adopt very different approaches to focus groups.

By contrast, focus groups can also have value to researchers who wish to compare divergent individual experiences to gain a greater understanding of a given phenomenon. Marketing researchers have long valued this feature of focus group research, because it reveals what aspects or experiences of a specific product appeal to a target population (Krueger and Casey 2009: 144). For academic researchers, this more phenomenological, individually focused approach can be particularly beneficial in the exploratory phases of research, as it can help researchers to generate useful hypotheses prior to gaining extensive background knowledge of a topic (Morgan 1988: 21). This is reflected within comparative policy research, where focus group methods have most commonly been employed in exploratory stages of research, and later complemented by other more confirmatory data sources (Dimitrova and Kortenska 2017; Gleeson et al. 2011; Hurrelmann 2015). Indeed, contrasting individual responses within a group setting can empower researchers to explore new ideas and causal explanations that they would not otherwise have pursued. In addition, by enabling individuals to share their perspectives within the group context, focus group research acts to empower participants such that they are more active contributors to the research project. This conforms to the aspirations of methodological approaches, such as community-based or participatory action research, which are increasingly popular in the social sciences, despite their relatively sparse use in fields such as comparative policy analysis.

The phenomenological epistemological perspective is more consistent with our approach to focus groups as a method for exploratory research. Despite the benefits we derived from observing processes of meaning construction among group participants, the most important takeaways of our focus group research were the trends that emerged when we aggregated discussion topics and responses. Specifically, during data analysis we noticed that participants often shifted focus from our central theme of social services, to emphasize medical services instead (Marier, Dickson, and Dubé 2017). Moreover, we observed that participants defined key concepts, such as their autonomy as service users, using language that was similar to the language used in assessment tools administered by social workers to determine service eligibility. These observations were exceptionally valuable because they gave us insight into how older adults define relevant policy problems related to their usage of public services. This not only gave us a greater descriptive edge when designing our survey instrument, but

also provided us with insight into problem definitions that will be instrumental to our larger comparison across cases.

5. METHODOLOGICAL DISADVANTAGES

The most apparent disadvantages of focus group research are related to the cost. Indeed, focus groups are both resource- and cost-intensive and therefore may not be worth the risk for many scholars, particularly in fields such as comparative public policy where the relative scarcity of existing studies leaves researchers with a lack of immediate referents to draw upon when creating their research designs. Cost considerations create pressure on researchers not to waste their resources on focus groups that fail as a result of short-sighted design or ineffective implementation. These types of failures can be the result of issues as seemingly benign as site selection, the time of day, or interpersonal dynamics between participants, and can result in biased and unusable data. Similarly, carelessness in the design of discussion questions, or the overzealous involvement by a moderator can also impart moderator-induced biases onto participant responses, which may act to artificially skew the data in the direction of researcher expectations (Stewart, Shamdasani, and Rook 2007: 85). Another important consideration pertains to the transcription and subsequent analysis of focus group data, which is made difficult by the potential cacophony of voices within the group discussion. While these problems can be mitigated either by video-recording or in-person observation by a member of the research team, this involves even more logistical planning. In this way, focus groups are resource-intensive both in terms of money (e.g. administration or travel costs), and time (e.g. staffed hours planning focus groups or analyzing the data), and may therefore be beyond the means of some researchers.

The uncommon use of focus group methods in comparative policy analysis can also present methodological disadvantages. As mentioned, the shortage of existing comparative policy studies that employ focus group methods leaves researchers without obvious referents from which to draw instruction. Although focus groups have been employed in policy analysis by powerful think tanks (Kahan 2001) and have been sparsely used by policy scholars to address comparative dynamics within single cases (Gamson 1992; Keeney, Von Winterfeldt, and Eppel 1990), they have been less commonly employed in comparative policy analysis across cases. This leaves scholars without a broad foundation of focus group research designs which they can adapt or replicate to address research questions relating to the implications of policy across cases. As such, in order to avoid undesirable outcomes such as group-think, inter-group conflict, or other forms of participant bias, particular attention must be paid to the design and implementation of focus group methods. In the following sections, we discuss how to avoid the potential disadvantages of focus group methods by describing best practices of focus group design and implementation, in turn.

6. DESIGNING FOCUS GROUPS

Using focus groups correctly in comparative policy analysis requires understanding the type of data that they are best suited for. As we have seen, there are pronounced differences between the individual-level data that are gleaned from interviews and the group-level data that are

produced by focus groups. Accounting for this difference has important implications for focus group design, which must anticipate the interaction effects that are likely to arise. Moreover, because focus groups are inherently artificial environments, there is no way to know if the type of interactions observed between participants would occur in a more natural setting (Green and Hart 1999). As such, it is necessary to meticulously design the focus group context, such that the data produced closely matches the research questions, and avoids potential pitfalls of focus group research. To this end, two considerations warrant particular attention: moderator involvement and group composition.

Moderator involvement is a vital dimension of focus group design because it has important implications for the group context. Morgan (1996) discusses moderator involvement along a continuum from more structured, where the moderator frequently intervenes with standardized, predetermined questions, to less structured focus groups where the moderator allows the participants significant autonomy in determining emergent topics of conversation. A key takeaway is that in order to capitalize on the value of focus groups in exploratory research, moderators must be willing to cede control of discussion topics to group members, allowing for emergent topics that can push the research in new directions. This can seem antithetical to scholarly understandings of research design, which value the researcher's ability to steer conversations towards a general topic of interest. Moreover, this carries the methodological disadvantage of wasting valuable time on discussions that are unrelated to primary main research questions – a cost can be later multiplied through the transcription and coding phases. Conversely, where moderators do assert more control over the discussion, they must contend with the potential for dishonesty or hesitancy with responses, which can result from relational dynamics present within the group context (Hollander 2004). This reflects a significant departure from the context of the structured individual interview, where these outcomes are both less likely and more easily diagnosable (Mosley 2013: 7). In this way, the focus group moderator must often cede control over both the topics of discussion and/or the presumption of honesty within the group context.

To mitigate the effects of these disadvantageous aspects of focus groups researchers must devote great energy to the composition of the groups themselves. In particular, they must ensure that participant selection matches the research question, such that conversation topics will be appropriate for all participants (Litosseliti 2003: 20). This is especially important when researchers aim to compare group perspectives on specific policy issues. This is exemplified by Price, Nir, and Cappella (2005) who compare different issue framing effects on opinions of same-sex civil unions across 50 online focus groups composed according to political ideology: conservative, liberal, and mixed. A research design such as this requires detailed knowledge of participants beforehand and effective recruitment strategies, which may be beyond the means of many researchers. Indeed, focus groups can be quite cost- and resource-intensive in the planning stage. However, getting group composition right is especially important in focus group research because it is much more difficult to organize and reconvene a group at a later date than it is with individual interviews (Kidd and Parshall 2000: 299). As such, the cost and effort required to effectively compose focus groups is a disadvantage of the method; however, these expenditures pay dividends when it comes to avoiding further pitfalls of focus group research that correspond to managing group dynamics.

7. MANAGING GROUP DYNAMICS

Focus groups are difficult to organize, and thus the potential value of the data must warrant selecting this over other approaches. Given that group interaction is the most valuable aspect of focus group data, researchers must prepare themselves to encourage positive and fruitful interaction amongst participants. In addition, the researcher must be skilled at managing the potentially difficult personalities of participants, as there is the tendency for social hierarchies and power dynamics to shape the quality and quantity of responses in certain contexts (Hertel, Singer, and Van Cott 2009: 307). Specifically, biases and group-think can be introduced quite easily to the focus group setting, particularly when negotiating the differences in aggressive and passive personalities in group conversations. Of the numerous characteristics that can affect group dynamics, many can be addressed in the design phase, including demographic factors such as age, gender, race and socio-economic status, and even physical characteristics such as size, health, and appearance. However, researchers often have no access to the personality characteristics of participants prior to the group session, and these are highly important group interactions, particularly when individuals with more aggressive personalities are included (Stewart, Shamdasani, and Rook 2007: 23). Moderators, therefore, must be adept at encouraging equitable discussion dynamics, and empowering more passive voices to engage and participate in the group context.

This is not to suggest that aggressive personalities should be avoided in focus group construction. Indeed, dynamics of both consensus and disagreement in the construction of group consciousness can yield important observations for the researcher by providing unique insight into why people form their opinions, and how strongly these opinions are held (Gamson 1992). This can involve broaching contested topics, and encouraging strong opinions. Nonetheless, moderators may have to assert control over the amount of discussion time allotted to particularly aggressive or loquacious participants because they can dramatically alter the responses of other respondents. Moderators must also be aware of manipulating respondents into conforming to preformed expectations regarding the subject matter. The introduction of homogenizing biases through manipulation or group-think makes it difficult to distinguish between individual views and group views when conducting data analysis (Litosseliti 2003: 20). These biases also act to significantly constrain variation in individual contributions, which is particularly detrimental to the use of focus groups in exploratory applications. To limit the homogenizing effects of moderator intervention, it is advisable to avoid structuring or steering the conversation as much as possible within the research design. Similarly, to minimize the focus on aggressive or dominant speakers, researchers can employ pre-testing strategies that allow them to identify problem participants, and subsequently to neutralize their negative impact through conversational strategies such as withholding eye contact or seating them closer to the moderator (Krueger and Casey 2009). Nonetheless, even these rather invasive strategies do not entirely eliminate the potentially negative effects of uneven group dynamics, which must be seen as a major disadvantage of focus group methods.

The group dynamics in our focus group research were consistently positive, and we had no experiences of either major conflict or total consensus among group participants. Nonetheless, within our groups there was often at least one dominant speaker among the participants. We found that effectively minimizing the participation of dominant speakers was particularly necessary when dealing with topics where there was disagreement among the group. In these instances, participants were especially engaged, and actively indicated a desire to share their

experiences. To facilitate their involvement, moderators found that the most effective strategy was to redirect the conversation from dominant speakers towards more passive participants using verbal prompts. However, we also found that asserting this kind of direct control over the conversation had a palpable effect on the flow of discussion and was thus best used sparingly as a last resort. In general, we tended to limit moderator involvement to the posing of questions and redirecting conversation away from dominant speakers or from unrelated discussion topics. Navigating the complexity of these dynamics makes focus groups particularly difficult to facilitate, and this is another drawback of the method. It is possible that environmental or group dynamics can significantly alter the social context, and consequently skew the data. In these situations, despite the time and energy that go into planning and facilitating the groups, it may be best to discard the compromised data, rather than aggregate it with data from the other groups. More generally, it is important for researchers to be intimately aware of the various disadvantages of focus group methods so that they can approach the data collection process with clear expectations. Armed with this awareness, they are more likely to benefit from the unique advantages that these methods can provide to suitable research designs.

8. CONCLUSION

Focus groups are rarely used in public policy, and even more rarely in comparative public policy, despite the fact that they are quite common in other social science disciplines. Interestingly, their usage is typically complementary to other methods such as surveys, individual interviews, and statistical analysis. The use of focus groups with other methods makes it quite difficult for the reader to gather specifically what is the added value of the focus group with regards to the content of the material gathered and the arguments deployed in research contributions. In this chapter, we described multiple ways in which focus groups can improve a research design while providing concrete examples from an ongoing project in social gerontology.

The focus group method provides unique benefits that can substantially improve a research project, such as the ability to provide a much deeper understanding of a policy issue than a simple survey and to allow for the identification of shared policy beliefs within a given group. To this end they also empower research participants to engage in interactive processes of meaning construction that reflect their specific shared experiences. This has the potential to embolden the perspectives of socially disadvantaged groups, whose concerns may be ignored in policy design and implementation. In addition, focus groups help to alleviate the strong reliance on policy experts in the field, and address part of the disconnect between the *makers* of public policy and its *users* (recipients, street-level bureaucrats, managers, etc.). At a time when the policy environment is increasingly complex, focus groups represent an excellent tool to capture how policies and programs are being experienced by diverse and interrelated publics.

NOTES

1. Three other articles made references to focus groups, but they were not a core element of the methodology. In the first case, the author relied on an extensive report from secondary sources that featured data from six focus groups (Moynihan 2008). In the second case, a footnote made reference to a focus group conducted by the author that did not feature in the methodology section (Blaydes

- 2014). Lastly, in an article on European citizenship, the author provided a quote and utilized some material from a previous study that relied on a focus group (Hurrelmann 2015).
2. This is a project financed by the Social Sciences and Humanities Research Council of Canada entitled The Politics of Social Gerontology (435-2014-1476).

REFERENCES

- Adler, G. and Rottunda, S. (2006). Older adults' perspectives on driving cessation. *Journal of Aging Studies* 20 (3): 227–35.
- Béland, D. (2005). Ideas and social policy: a historical institutionalist perspective. *Social Policy and Administration* 39 (1): 1–18.
- Blaydes, L. (2014). How does Islamist local governance affect the lives of women? *Governance* 27 (3): 489–509.
- Bowling, A. (2007). Aspirations for older age in the 21st century: what is successful aging? *International Journal of Aging and Human Development* 64 (3): 263–97.
- Caillaud, S. and Flick, U. (2017). Focus groups in triangulation contexts. In R. S. Barbour and D. L. Morgan (Eds.), *A New Era in Focus Group Research*. London: Palgrave Macmillan, 155–77.
- Carey, M. A. (1994). The group effect in focus groups: planning, implementing, and interpreting focus group research. In J. Morse (Ed.), *Critical Issues in Qualitative Research Methods*. Thousand Oaks, CA: Sage, 225–41.
- Chrisler, J. C., Barney, A., and Palatino, B. (2016). Ageism can be hazardous to women's health: ageism, sexism, and stereotypes of older women in the healthcare system. *Journal of Social Issues* 72 (1): 86–104.
- Claes, R. and Heymans, M. (2008). HR professionals' views on work motivation and retention of older workers: a focus group study. *Career Development International* 13 (2): 95–111.
- Couture, M., Israël, S., Soulières, M., and Sasseville, M. (2019). Implementing a systematic screening procedure for older adult mistreatment within individual clinical supervision: is it feasible? *Journal of Interpersonal Violence* 34 (13): 2813–33.
- Dimitrova, A. and Kortenska, E. (2017). What do citizens want? And why does it matter? Discourses among citizens as opportunities and constraints for EU enlargement. *Journal of European Public Policy* 24 (2): 259–77.
- Emler, C. A. and Mocerri, J. T. (2012). The importance of social connectedness in building age-friendly communities. *Journal of Aging Research*. <http://dx.doi.org/10.1155/2012/173247>.
- Estes, C. L. and Binney, E. A. (1989). The biomedicalization of aging: dangers and dilemmas. *The Gerontologist* 29 (5): 587–96.
- Gamson, W. A. (1992). *Talking Politics*. Cambridge: Cambridge University Press.
- Gerring, J. (2001). *Social Science Methodology: A Criterial Framework*. Cambridge: Cambridge University Press.
- Gleeson, D., Legge, D., O'Neill, D., and Pfeffer, M. (2011). Negotiating tensions in developing organizational policy capacity: comparative lessons to be drawn. *Journal of Comparative Policy Analysis: Research and Practice* 13 (3): 237–63.
- Green, J. and Hart, L. (1999). The impact of context on data. In R. S. Barbour and J. Kitzinger (Eds.), *Developing Focus Group Research: Politics, Theory and Practice*. London: Sage, 21–35.
- Hagen, J. L. (1994). Women on welfare talk about reform. *Public Welfare* 52 (3): 30–42.
- Hecl, H. H. (1972). Policy analysis. *British Journal of Political Science* 2 (1): 83–108.
- Hertel, S., Singer, M. M., and Van Cott, D. L. (2009). Field research in developing countries: hitting the road running. *PS: Political Science & Politics* 42 (2): 305–9.
- Higgs, P., Hyde, M., Wiggins, R., and Blane, D. (2003). Researching quality of life in early old age: the importance of the sociological dimension. *Social Policy & Administration* 37 (3): 239–52.
- Hollander, J. A. (2004). The social contexts of focus groups. *Journal of Contemporary Ethnography* 33 (5): 602–37.
- Hurrelmann, A. (2015). Demoi-cratic citizenship in Europe: an impossible ideal? *Journal of European Public Policy* 22 (1): 19–36.

- Hyde, M., Wiggins, R. D., Higgs, P., and Blane, D. B. (2003). A measure of quality of life in early old age: the theory, development and properties of a needs satisfaction model (CASP-19). *Aging & Mental Health* 7 (3): 186–94.
- Kahan, J. P. (2001). Focus groups as a tool for policy analysis. *Analyses of Social Issues and Public Policy* 1 (1): 129–46.
- Kaufman, S. R. (2004). Revisiting the biomedicalization of aging: clinical trends and ethical challenges. *The Gerontologist* 44 (6): 731–8.
- Keeney, R. L., Von Winterfeldt, D., and Eppel, T. (1990). Eliciting public values for complex policy decisions. *Management Science* 36 (9): 1011–30.
- Kidd, P. S. and Parshall, M. B. (2000). Getting the focus and the group: enhancing analytical rigor in focus group research. *Qualitative Health Research* 10 (3): 293–308.
- Knodel, J. (1995). Focus groups as a qualitative method for crosscultural research in social gerontology. *Journal of Cross-Cultural Gerontology* 10 (1): 7–20.
- Koch, T. and Webb, C. (1996). The biomedical construction of ageing: implications for nursing care of older people. *Journal of Advanced Nursing* 23 (5): 954–8.
- Koenig-Archibugi, M. and Macdonald, K. (2013). Accountability-by-proxy in transnational non-state governance. *Governance* 26 (3): 499–522.
- Krekula, C. (2007). The intersection of age and gender: reworking gender theory and social gerontology. *Current Sociology* 55 (2): 155–71.
- Krueger, R. A. and Casey, M. A. (2009). *Focus Groups: A Practical Guide for Applied Research*. Thousand Oaks, CA: Sage.
- Lancaster, K., Santana, L., Madden, A., and Ritter, A. (2015). Stigma and subjectivities: examining the textured relationship between lived experience and opinions about drug policy among people who inject drugs. *Drugs: Education, Prevention and Policy* 22 (3): 224–31.
- Litosseliti, L. (2003). *Using Focus Groups in Research*. London: Bloomsbury.
- Marier, P., Dickson, D., and Dubé, A.-S. (2017). Autonomy in context: perspectives from older adults in the province of Quebec. Presentation at the Canadian Association on Gerontology 46th Annual Scientific & Educational Meeting, Winnipeg, October 19–21.
- McKeown, B. and Thomas, D. B. (2013). *Q Methodology* (2nd edition). London: Sage.
- Merton, R. K., Fiske, M., and Kendall, P. A. (1990 [1956]). *The Focused Interview: A Manual of Problems and Procedures*. New York: Free Press.
- Morgan, D. L. (1988). *Focus Groups as Qualitative Research*. Thousand Oaks, CA: Sage.
- Morgan, D. L. (1996). Focus groups. *Annual Review of Sociology* 22 (1): 129–52.
- Mosley, L. (2013). Introduction: “Just talk to people”? Interviews in contemporary political science. In L. Mosley (Ed.), *Interview Research in Political Science*. Ithaca, NY: Cornell University Press, 1–28.
- Moynihan, D. P. (2008). Combining structural forms in the search for policy tools: incident command systems in US crisis management. *Governance* 21 (2): 205–29.
- Orel, N. A. (2004). Gay, lesbian, and bisexual elders: expressed needs and concerns across focus groups. *Journal of Gerontological Social Work* 43 (2–3): 57–77.
- Pierson, P. (1993). When effect becomes cause: policy feedback and political change. *World Politics* 45 (4): 595–628.
- Price, V., Nir, L., and Cappella, J. N. (2005). Framing public discussion of gay civil unions. *Public Opinion Quarterly* 69 (2): 179–212.
- Rogers, R. R. (2013). Using interviews to understand racial group identity and political behavior. In L. Mosley (Ed.), *Interview Research in Political Science*. Ithaca, NY: Cornell University Press, 225–43.
- São José, J. M. S., Amado, C. A. F., Ilinca, S., Buttigieg, S. C., and Larsson, A. T. (2019). Ageism in health care: a systematic review of operational definitions and inductive conceptualizations. *The Gerontologist* 59 (2): e98–e108.
- Schneider, A. and Ingram, H. (1993). Social construction of target populations: implications for politics and policy. *American Political Science Review* 87 (2): 334–47.
- Stapenhurst, R. and Pelizzo, R. (2012). Improving democracy and accountability in Ghana: the importance of parliamentary oversight tools. *Governance* 25 (2): 335–46.
- Stewart, D. W. and Shamdasani, P. N. (2015). *Focus Groups: Theory and Practice* (3rd edition). Thousand Oaks, CA: Sage.

- Stewart, D. W., Shamdasani, P. N., and Rook, D. W. (2007). *Focus Groups: Theory and Practice* (2nd edition). Thousand Oaks, CA: Sage.
- Sussman, T., Kaasalainen, S., Bui, M., Akhtar-Danesh, N., Mintzberg, S., and Strachan, P. (2017). "Now I don't have to guess": using pamphlets to encourage residents and families/friends to engage in advance care planning in long-term care. *Gerontology and Geriatric Medicine* 3. <https://doi.org/10.1177/2333721417747323>.
- Walsh, K. and O'Shea, E. (2008). Responding to rural social care needs: older people empowering themselves, others and their community. *Health & Place* 14 (4): 795–805.
- Welford, C., Murphy, K., Rodgers, V., and Frauenlob, T. (2012). Autonomy for older people in residential care: a selective literature review. *International Journal of Older People Nursing* 7 (1): 65–9.
- Wiles, J. L., Leibing, A., Guberman, N., Reeve, J., and Allen, R. E. (2012). The meaning of "aging in place" to older people. *The Gerontologist* 52 (3): 357–66.
- Wilkinson, S. (1998a). Focus groups in feminist research: power, interaction, and the co-construction of meaning. *Women's Studies International Forum* 21 (1): 111–25.
- Wilkinson, S. (1998b). Focus group methodology: a review. *International Journal of Social Research Methodology* 1 (3): 181–203.
- Wutich, A., Lant, T., White, D. D., Larson, K. L., and Gartin, M. (2010). Comparing focus group and individual responses on sensitive topics: a study of water decision makers in a desert city. *Field Methods* 22 (1): 88–110.
- Yassuda, M. S., Wilson, J. J., and Mering, O. V. (1997). Driving cessation: the perspective of senior drivers. *Educational Gerontology: An International Quarterly* 23 (6): 525–38.

18. Using ethnography in comparative policy analysis: premises, promises and perils

Raul Pacheco-Vega

1. INTRODUCTION

Undertaking comparative public policy analysis requires us to think broadly about types of policies we are required to analyze, tools we should use to investigate them and strategies to implement research projects and policy analytical programs. While qualitative methods have a very broad repertoire of tools to choose from, it is important to ensure that we choose the right approach for the public policy issue we are researching, as well as the appropriate implementation strategy for the method we've chosen. In this chapter, I focus on ethnography as a robust research method to study comparative policy analysis.

While there are numerous other approaches to studying and engaging in policy analysis from a comparative perspective, ranging from qualitative to quantitative to spatial, ethnographic research strategies can yield insights we would not be able to gain through other methods. Ethnography enables a researcher to embed him/herself in a specific community in a way that can provide in-depth coverage and analysis of policy issues that are often rendered invisible if we use other methods, even qualitative ones. Frequently, the only way to properly uncover thorny issues with public policy implications is undertaking in-depth ethnographic fieldwork. This demonstrates the importance of including ethnography in the repertoire of research strategies available to comparative policy analysts.

In this chapter, I engage with a broad range of literature, though I narrow down my literature review by focusing on scholarly works that have both public policy implications and a focus on ethnography as a research method. Moreover, I use a comparative lens to explore the literature. While single-case studies can have insights for comparative policy analysis, here I follow Wolf and Baehler's lead, drawing lessons that can be applicable to various cases despite using a single unit of analysis, but looking toward using these lessons, insights and learnings and applying them to larger N studies, even if N is still within the medium range (three to six cases).

The chapter is organized as follows: in the second section, right after this introduction, I outline what is ethnography, how and when we use ethnographic methods in policy research, and what limitations the method has. In this section, I also outline how ethnography works, and how we can link comparative methods with ethnography more clearly. I take a three-pronged approach to explaining how to use ethnography in comparative policy analysis. First, I provide an overview of the method, and explain how it could be used to study comparative public policy, and to undertake comparative policy analysis. Second, I detail how comparative ethnographies can aid policy analysis. And third, I explain how single-case ethnographies can help us draw insights for comparative policy analysis. It would seem like a waste to consider single-case ethnographies when we are writing about comparative work. However, as scholars of the comparative method in political science and policy studies would argue, even

single-case studies can provide insight into how comparison works, and how to undertake structured, focused cross-case analyses.

While several of the ethnographies cited in this chapter are specific to one case study, there are very clear and tangible ways to cross-link comparative methods with ethnography and policy analysis. I lay out the contributions that the method itself can provide and how we can link methodology usage with policy analytical goals. In the third section, I outline how the method can be applied in comparative policy analysis, offer a summary of each one of the stages that my framework outlines, and provide a number of examples of how comparative policy analytical lessons can be drawn from various ethnographies.

2. WHAT IS ETHNOGRAPHY AND HOW CAN IT BE APPLIED IN COMPARATIVE POLICY ANALYSIS?

Ethnography as a research method involves engaging in deep observation of a phenomenon, as well as embedding oneself in the site where the research is taking place. Ethnographic work seeks to understand (much like the root of its method in the anthropological sciences) cultures (McGranahan 2014). While there's a discussion on whether fieldwork is necessarily anthropological and thus ethnographic, McGranahan argues that it is possible to teach ethnographic sensibility without having to necessarily enter a research site. What McGranahan suggests in her teaching-oriented piece is that we ought to teach students (and by extension, I would argue, policy analysts) to consider the broad variety of cultural contexts, practices, and individual characteristics that make up a research site.

As McGranahan writes:

Ethnography is the writing of the people, the writing of society, the writing of culture. Ethnographies have long been what anthropologists write and read, but recently we have also been using the term as a shorthand for fieldwork, saying we are “doing ethnography” when we mean ethnographic research. By ethnographic research, anthropologists mean the ever-evolving Malinowskian program of an ethnographer in the field conducting participant-observation paired with a range of other methods, living within a community, and getting deeply into the rhythms, logics, and complications of life as lived by a people in a place, or perhaps by peoples in places. Ethnographic research, then, is more than a method. (McGranahan 2014: 23)

While there are dozens of methodological book manuscripts that have been published on the topic of ethnography and how to undertake an in-depth, fieldwork-based study using ethnographic approaches, I will not survey the vast literature on the topic in this chapter. However, I do want to help the reader situate him/herself in the broader context of why it is that we use ethnography in comparative policy analysis and where in the analytical process methodological choices are made. Fundamentally, ethnography is about culture, about understanding how communities, individuals, and societies live. Ethnography also asks whether and how their lived experiences have shaped choices, decision-making processes, and individual behaviors.

One of the key elements of ethnography as a research method, and definitely the marker of distinction with regards to other methodologies and research paradigms is the explicit focus on culture. Ethnography is different from elite interviews and other methods that are often used in public policy analysis, public administration, political science, and public management in that it focuses on the specific elements of culture that are not captured by other research

methods. We use ethnography, the deep embedding of a researcher within a policy domain, a community, or an organization, in order to discern the culture within specific societal branches (Clifford and Marcus 1986). Interviews can often simply be brief and have a much narrower temporal horizon (Mosley 2013), whereas ethnography requires deeply engaging with the community and/or research subject under study. Ethnography, therefore, is rather suitable for issues that public policy scholars and particularly comparative policy analysts would be interested in analyzing. It is clear as well from the way in which the method works that undertaking ethnographic work will require substantial time and energy investment on the part of scholars. This long-ranging, standing commitment to studying a specific community, agency, or organization may or may not be rewarded within our current scholarly framework, given the pressure to continuously publish journal articles and book chapters. Nevertheless, there is enormous value in investing in ethnographic research training of policy scholars and further exploration of policy problems directly on the field.

Though ethnography is used in a broad variety of disciplines and fields of scholarship, some of the most interesting applications include museum ethnography (Burt 1998) and corporate (organizational) ethnography (Ladner 2014). Traditionally, human geographers, sociologists, and (less frequently so) political scientists borrowed from anthropology and applied ethnographic approaches to understand key issues. But as recent surveys of the field have shown (Becker et al. 2004; Cappellaro 2017; Huby, Harries, and Grant 2011), there's been a renewed interest in applications of ethnographic approaches in the policy sciences. Moreover, as Pacheco-Vega and Parizeau have shown, ethnography can provide a useful methodological framework to engage with vulnerable communities in a way that is meaningful and also non-extractive by doing what they call a doubly-engaged ethnography (Pacheco-Vega and Parizeau 2018).

As early as 1981, Rist had already called attention to the value of ethnography as a research method. His main concern was that the policy sciences were veering toward a more complex and confusing model of undertaking studies of policy design, implementation, and evaluation (Rist 1981). From problem definition to instrument choice, ethnographic engagement with local communities facilitates learning from those individuals whose lives will be affected by decision-making processes from which they are often excluded. As Rist outlines, “[q]ualitative research can contribute by means of restricting the problem definition, by isolating the levers of change, and by identifying unintended consequences of policy decisions” (Rist 1981: 487).

Van Hulst also showed how ethnographers can make sense of how local governments' decision-making processes have an impact on communities. As van Hulst indicates:

[e]thnographic work brings something special to the study of sense-making in local governance: the ethnographer's access to the experiences lived by the people under study. In addition, ethnographers not only look for the experiences of the people in and around local government, they also draw on their own experiences. Because the experiences of politicians, administrators, bureaucrats, professionals and citizens are both the result of and the basis for their acts, understanding these experiences helps ethnographers to explain the practice of local governance. (van Hulst 2008: 143)

Van Hulst makes explicit the reasons why ethnographic methods offer great contributions to the study and practice of comparative policy analysis: they allow researchers to understand all stakeholders' viewpoints. It is much harder to gain these understandings through a simplistic quantitative analysis. Even qualitative semi-structured interviews are not able to provide as much in-depth insight as deeply embedded fieldwork. While van Hulst's description is

specific to one field site, engaging in comparative analysis isn't hard because all we really need to do is simply increase the number of case studies or leave one constant and vary other dimensions, even within the same case study.¹

We can find ethnography used as a method in a very broad range of policy issue areas, from language policy (Johnson 2009), foreign policy (Kuus 2013), and welfare control and redistributive policies (Dubois 2009) to health policy (Erasmus 2014; Walt et al. 2008), urban regeneration (Crossa 2012; Davies 2002; Guarneros-Meza 2009; Mah 2010), and citizen participation strategies in comparative urban policy (Maginn 2006). But as I will outline in further sections of this chapter, there has always been a reciprocal, symbiotic, and intimate relationship between anthropology and public policy.

Hackenberg indicated that Bohannan first argued how anthropology was supposed to engage with the policy sciences by providing a compass and directional strategy to focus on issues that are relevant to society rather than abstract, simplistic examinations of public agencies, bureaucracies, and political life (Hackenberg 1985). Around the early 1980s, conversations between anthropology and public policy began in earnest, although it wasn't until later that policy scientists began taking anthropology more seriously and engaging with the literature. More importantly, a new wave of interest in ethnographic, intense, time-consuming, in-depth fieldwork-based approaches for policy studies began around the mid-1980s to early 1990s. Obviously, this coincided with a search for policy relevance, a phenomenon that has also regained traction in the political science literature. Even experimental political science is beginning to look for ways to integrate qualitative, ethnographic approaches with field experiments (Paluck 2010).

3. ETHNOGRAPHY IN PUBLIC POLICY, PUBLIC MANAGEMENT, AND PUBLIC ADMINISTRATION RESEARCH²

Most scholars of public policy will probably remember Dvora Yanow's *Conducting Interpretive Policy Analysis*, perhaps the most cited single-author book on policy analysis using qualitative methods. Yanow (2000) has almost single-handedly taught two generations of policy analysts how to conduct this type of work, even if her book isn't specific to comparative analysis.

One can trace the growing interest in applying anthropological research methods to studies of public policy to a few works by Shore, Wright, Seidel, Vidal, Wedel, Feldman, and several other scholars. The anthropology of public policy emerged as a new field of study because this discipline is well situated to understand the complexity and messiness of public policy making processes (Wedel and Feldman 2005; Wedel et al. 2005). While there's much work done on social policy (Okongwu and Mencher 2000), this renewed interest in an anthropological examination of policy processes, mechanisms, actors, and outcomes has sparked important innovations in how we study public policies across a broad range of issue areas.

While there's some literature on the use of ethnographic methods in corporate governance (Trondman 2000), and public management (Cappellaro 2017; Huby, Harries, and Grant 2011), there's still a dearth of reliable sources on the application of ethnography to undertake comparative policy analysis. Ethnography in comparative policy analysis has been relatively shunned, as Geva-May and coauthors have shown in a recent analysis of 20 years of *Journal of Comparative Policy Analysis (JCPA)* articles. Only 2% of all articles surveyed by Geva-May

et al. are reported as having an ethnographic methodological strategy. This means that just 7 pieces published in the last 20 years of *JCPA* issues explicitly use ethnography as a method for comparative policy analytical work (Geva-May, Hoffman, and Muhleisen 2018). Contrast this figure with 55 pieces that use quantitative approaches (15%) and 219 case-based (62%). While *JCPA* is not as quantitative as, say, the *Journal of Public Administration Research and Theory* (*JPART*), it is quite clear that ethnography does not enjoy the same degree of popularity that other methodological strategies have. In this section, I survey available literature and examine different approaches to the application of ethnography in comparative policy analysis.

Naturally, one can safely assume that case study research can also involve ethnographic approaches, but it is striking that comparative policy analysis does not seem to have an appreciation for the benefits that in-depth observation over a substantially long period of time can have. In this chapter I answer the question: what can ethnography contribute to our understanding of comparative policy analysis and how can insights from ethnographic research methodologies be best implemented in policy analysis of comparative cases? While there is much value in presenting specific, detailed cases of comparative policy analysis where ethnography is used to learn from the community under study, throughout this chapter I take a much more focused approach. Instead of presenting a detailed case study of a comparative policy analysis ethnography, I draw lessons from a broad range of scholarly works and offer an analytical examination of the utility of the method itself. While I do survey literature across disciplines and policy issue areas, I maintain a narrow focus on how the method addresses the policy problem under study.

Throughout the chapter, I emphasize a variety of methodological challenges that comparative policy research problems can present, how theory can be applied to specific issues, and what the scope of comparison is. It's important to note that while ethnographic work can be reported at the national level, by its very nature it is a method that is more amenable to sub-national comparisons, particularly across communities in one country, state, or metropolitan area. One could also potentially conduct comparative ethnographic work across different states and countries, but the generalizability of those findings could be challenged simply because of the sheer amount of work and number of cases that would be required to properly and reasonably report national-level patterns. Finally, while it is clear ethnographic work lends itself to cross-scalar comparisons (from the neighborhood level to the local scale to the federal), insights don't necessarily "scale". One should be wary of, for example, generalizing about Canadian federal-level policies and their impacts on local communities when drawing from case studies of specific locations.

Comparison is inherent to the study of policy analysis and its pragmatic application to researching and providing solutions to public issues that are deemed important and appear on governmental agendas. As Wolf and Baehler aptly state:

Comparison lies at the heart of all policy analysis. Initially, policy students may learn how to compare different policy interventions in theory and in application to a specific problem. Analysing "what should be done" to improve situations involves applying specialist techniques to project (based on trends), predict (based on theory) and conjecture (based on expert judgment). (Wolf and Baehler 2018: 420)

Comparing is inherent to both ethnography and comparative policy analysis (Vogtz 2002). This point is brought home by Vogtz quite masterfully as she conducts a multi-case study with fieldwork in four primary schools in Switzerland and England. She used a variation of a 2×2

matrix (English Large, English Small, Swiss Large, and Swiss Small) to choose the types of schools where she would be engaging in fieldwork. She categorized these institutions by size and country. A multi-site ethnography of this sort can be easily used to undertake comparative education policy analysis, which in the end was the main goal that Vogtz pursued.

However, not every ethnographic comparison must be multi-country, cross-national. A researcher could easily draw lessons from single-country, sub-national cross-unit comparisons. Even more so, one should not be wary of using single-case studies, particularly because iteration can enable researchers to generate enough understanding about a phenomenon that it becomes easier to analyze even within the confines of a single case study. I quote Wolf and Baehler again:

Specifically, we suggest that an essential aspect of lesson drawing – gaining a new, plausible idea – draws on policy professionals’ natural capability to iterate between cases that supply transferable lessons and cases to which those lessons may be best applied. (Wolf and Baehler 2018: 420)

Methodologically speaking, Wolf and Baehler provide a very useful template for how we should conduct comparative policy analysis, using a single-case study as the cornerstone of our examination. While their discussion is specific to policy transfer, it can be easily translated to the application of ethnographic approaches to lesson drawing. In this case:

For the policy professional at the centre of a transfer, attention is essentially comparative: a receptive and attuned learner acquires and judges plausible lessons from B for application in A to achieve A. (Wolf and Baehler 2018: 421)

Finally, Wolf and Baehler identify the crux of any attempt to use ethnography as a research strategy for comparative policy analysis. As they indicate, the research process (and by extension, the analysis) is iterative, and dialogic. Analysts should engage in repeated dialogue with communities under study and with stakeholders across all levels and sectors. Following Wolf and Baehler:

Comparative case learning can be enhanced through the dialogic interaction in network exchanges between the source and target case actors, the memories of the past and prospective mages that are presented, triggered and created in the course of those interactions. (Wolf and Baehler 2018: 421)

4. THREE MODES OF ETHNOGRAPHIC ENGAGEMENT IN COMPARATIVE POLICY ANALYSIS

There are at least two main modes of inquiry in comparative policy analysis using ethnographic methods. One way is purposefully choosing ethnography as a research method and creating an inquiry/investigative strategy around the project so that researchers embed themselves in field sites from the beginning. I call this a “push approach”. This approach is purposive, intentional, crafted, because here the methodological choice is made before engaging in comparative policy analysis. This strategy “pushes” ethnography to the core of the project and demands from researchers that they think about case study selection, tactics for engagement with community members, and processes to enter field sites as well as exit strategies. The “push approach” makes comparative policy analysis the core framework of thinking and eth-

nography the fundamental mode of scholarly and analytical inquiry. From a methodological perspective, using a “push approach” necessitates a lot of previous work before engaging in fieldwork, or even thinking about case study selection. This method, however, is much more robust in terms of ethics and also logistically is the one that makes most sense if policy issue areas are thorny and require a lot of care in developing a strategy for fieldwork deployment. There are various subjects that can be this worrisome, like border policy, security policy, drug policy, and homelessness policy.

Another way of using ethnography in comparative policy analysis is to draw from already-published ethnographic accounts and systematically examining their content, looking for patterns, processes, and ideas that can be applied to policy analytical work in comparative perspective. I call this the “pull approach”. The “pull approach” focuses on exploring a broad range of ethnographic works to then draw insights from each one of the case studies and policy areas where this method has been applied and then, afterwards, synthesizing these learnings in a comparative policy analytical report. This methodological framework puts the issue area at the core, and therefore the ethnographic application is the main substantial data contributor. Comparative policy analysis then functions as a lens through which various ethnographies can be examined and lessons drawn. In this chapter I demonstrate the usefulness of the “pull approach”, albeit I also draw attention to potential concerns with the use of individual ethnographies developed in different contexts, where insights drawn may or may not be applicable in other areas or regions of study.

Moreover, there’s a third way of integrating ethnographic approaches to comparative policy analysis. I call this the “mixed approach”. In a mixed approach, we draw insights from the pull approach to better design potential fieldwork and choose case study sites more systematically and carefully. This strategy combines the best of both worlds, as it engages in an iterative process that can then be improved through repeated engagement, review, and revision.

To conduct this component of the literature review for this chapter, I read and analyzed 20 single-authored book-length volumes that are reported as ethnographies, in a broad range of policy issue areas and geographical regions, including housing in Milwaukee (Desmond’s *Evicted*, 2016), policing in Pennsylvania (Goffman’s *On The Run*, 2014), conservation in the Brazilian Amazonia (Kawa’s *Amazonia in the Anthropocene*, 2016), migration on the US–Mexico border (De Leon’s *The Land of Open Graves*, 2015), and foreign aid in Sierra Leone for HIV prevention programs (Benton’s *HIV Exceptionalism*, 2015). A few ethnographies focused on similar issues across different geographical scales (Millar’s 2018 *Reclaiming the Discarded* study of waste pickers in Rio de Janeiro, in Brazil, and Reno’s 2016 *Waste Away* analysis of waste picking in a Michigan landfill, in the US) and types of work (Robin Nagle’s *Picking Up*, 2014). I also examined other sole-authored, book-length ethnographic works that covered important issues that may be under-researched but have relevant policy implications, like Mears’s 2011 ethnographic account of the modeling industry.

Given that so many of these volumes were specifically single-case ethnographies, I also sought to integrate into my analysis cross-comparative case studies, though this proved quite difficult to do because of how much less popularity ethnography has for comparative policy analysis. Nevertheless, I found two volumes where this systematic comparative ethnography approach had a very specific comparative policy analysis focus. The first one is a classic of public administration: Robert Dahl’s *Who Governs?*,³ where Dahl undertook an ethnography of New Haven bureau chiefs, specifically discerning the role of political power and representation in pushing new policy reforms in an American city (Dahl 2005).

The second one is a comparative, cross-national ethnography of security policy in three cities across two continents, Laura Huey's *Negotiating Demands* (2007). Huey examines policing of skid row areas and their associated policies across San Francisco (California, USA), Edinburgh (Scotland, United Kingdom), and Vancouver (British Columbia, Canada). Using a similar approach to Goffman's *On the Run* and Forrest D. Stuart's *Down, Out, and Under Arrest* (2016), Huey conducts a cross-national, comparative policing study that yields insights into how the contextual factors of each country's regulatory framework and the domestic politics of each city's social fabric created different conditions for success or failure of security policies in marginalized areas (Huey 2007). Skid row policing is often a politically contentious topic for public administrators because of the specific conditions of the communities being monitored and secured. This is also one policy where those who oversee ensuring compliance with regulations and maintaining rule of law can often trespass their regulatory domains. As Stuart aptly asks, "who polices the police?" (Stuart 2016). This is an important question that cannot be easily addressed with traditional research methods, as it requires frequent and systematic observation of routine behaviors, often deeply embedded within a community or the police force itself.

While the single-case ethnographies volumes were not specifically written by public policy scholars, they provide important comparative public policy and policy analytical insights, even if in many of these cases authors conducted their ethnographies in a single country/city. To be able to draw lessons for comparative policy analysis from published works that weren't specifically written by policy scholars, one needs to look for insights that are specific to the policy sciences: how did conservation work in the Amazonian region and what lessons can we draw from ethnographic work for instrument design and implementation? Why isn't the modeling and fashion industry properly regulated and why isn't this regulation apparently or visibly part of any government's agenda-setting process?

One of my contributions with this chapter, as I demonstrate in the next sections, resides in making the connections between how ethnographies are conducted and whether they have explicit policy goals. Moreover, I highlight the comparative component and show how policy analysts can use these ethnographies to help bureaucrats and politicians make more informed policy decisions. I should note that I will not attempt to synthesize every one of the 20 volumes I've analyzed. Nevertheless, I will use a few vignettes from these published works to illustrate how their ethnographic work facilitates comparative policy analysis. Moreover, in doing an in-depth, broad-ranging, and far-reaching review of the literature, I also came across several journal articles that can illuminate how we think about analyzing policy using a comparative lens.

On comparative immigration policy and borderland security policy, we can draw many insights from Francisco Cantú's *The Line Becomes a River*. Cantú's experience as a border patrol agent illuminated his ethnographic study of the failures and shortcomings of border policy across the US–Mexico border (Cantú 2018). While we can use the United States as a case study of draconian border protection and security policies, Cantú demonstrates the variegated trajectories that immigration policies take when dealing with a broad range of heterogeneous populations. The US has extraordinarily tough border policies, but Cantú's work allows us to use a comparative lens to explore and attempt to understand the differences between US–Mexico and Mexico–Belize/Mexico–Guatemala border policies. While the US has extreme vetting processes, Mexico does not lag far behind and has possibly equally bad or even worse policies to deal with illegal immigrants.

Sectors that have been poorly regulated include the modeling and fashion industry and the meat commodity chain. Ethnographic work that investigates how these industries operate enables comparative policy analysts to draw attention to key regulatory issues, including discretionary government action, unregulated sectors, poor enforcement, and weak policy design. Both *Pricing Beauty* (Mears 2011) and *Every Twelve Seconds* (Pachirat 2013) offer insightful overviews of the key regulatory issues plaguing both modeling and fashion and meat production. While both ethnographies were conducted in the United States, insights drawn from both can be generalized to other countries' industrial performances. This demonstrates the potential for cross-national comparative policy analytical work.

On immigration policy, both De Leon's *Land of Open Graves* and Cantú's *The Line Becomes a River* offer gripping and painful ethnographies of border policy issues (Cantú 2018; De Leon 2015) across the US–Mexico border. Methodologically speaking, the ethnographic approach Cantú takes is quite different from De Leon's, as Cantú embodied policy and became a border security guard in order to conduct his in-depth fieldwork. De Leon uses ethnography, archaeology, linguistics, and forensic science, combining all these disciplines and methods masterfully. What makes these two volumes useful for comparative policy analysis is that they showcase different ways in which ethnography can be applied to understand a policy issue that affects not only the southern border of the United States, but by extension what's happening along the southern border of Mexico. Wolf's *Mano Dura* (2017) complements these texts by offering an examination of street gangs in El Salvador. Though El Salvador does not directly border Mexico, it does share a border with Guatemala, and therefore, many of the migratory issues that affect and drive US immigration policy respond by extension to violence in El Salvador. Even if not her main research goal, Wolf's work offers some insight into policy initiatives such as the Frontera Sur program and the effect of US immigration deterrent practices on immigration from El Salvador as well as a comparison of US and Salvadorean approaches to gang policy (another cross-national issue) (Wolf 2017).

Comparing water governance and infrastructure policy across countries is important if we want to draw relevant policy learnings and implement appropriate instruments across countries, regions, and cities. Nikhil Anand's *Hydraulic City* provides a much-needed examination of the various ways in which urban water policy in Mumbai fails, and how we can learn from those failures (Anand 2017). Throughout his book, Anand illuminates the multiple mechanisms by which infrastructure and water and sanitation delivery in the Indian city of Mumbai do not perform to the required standard. Moreover, Anand shows that the Mumbai case can also shed light on how informal water markets emerge to ensure robust service provision. It's important to note that other scholars have also shown similar patterns of emergence of informality in potable water and sanitation service provision in different cities in India, as Ranganathan shows in her study of water mafias in Bangalore (Ranganathan 2014, 2018).

Hoover's multi-disciplinary examination of activism against toxic emissions in a Mohawk community allows us to understand how politics and racism can play a prominent role in where Superfund sites are located. Environmental racism is prevalent in the United States and has been widely documented elsewhere, but where Hoover's ethnographic contribution can help comparative toxics policy is in synthesizing knowledge and information not only from traditional sources including scientists, experts, and activists, but also from the Indigenous communities where her ethnographic engagement occurred, in the community of Akwesasne in upstate New York (Hoover 2017). Hoover's work can also be linked to that of Sarah Ann Wylie and her ethnographic study of activism against fracking (Wylie 2018). Wylie's work

in comparison with Hoover's can offer enormous insight into how different communities respond to toxic emissions and polluting industrial processes occurring within the scope of their daily lives. A related account of communities affected by pollution in China by Anna Lora-Wainwright suggests again potential for comparative toxics policy analytic work, as one could compare community responses in three Chinese communities (Lora-Wainwright 2017) with how other countries' populations responded.

Global health issues affect domestic populations and offer great potential for comparative policy analytical work. Adia Benton's *HIV Exceptionalism* is one important example that cross-links the global sphere of action of the World Health Organization and its HIV programs and the domestic application of these approaches in Sierra Leone (Benton 2015). As Benton indicates:

Sierra Leone – like most African countries – benefits from a large influx of foreign aid money specifically targeted at HIV/AIDS prevention and care programs. This foreign aid comes, in large part, because international donors believe HIV is an exceptional condition, requiring a focused, intensive response that is unlike any directed to other diseases. (Benton 2015: x)

As I have shown in this section, even if ethnographies aren't written with a comparative policy analytical framework in mind, a pull approach allows us to combine insights from a broad range of disciplines, policy issues, and sectors. For example, one can just as easily gain insights for global environmental policies by examining conservation at sub-national scales, as Nick Kawa's *Amazonia in the Anthropocene* does (Kawa 2016). His examinations of pre-Columbian Amerindians and contemporary rural Amazonians have shaped conservation policies. Using a cross-temporal approach, Kawa's work allows me to show by extension how comparative policy analytical work can be undertaken within the same geographical region at two points in time. Moreover, as Benton and Kawa's ethnographies offer bottom-up (Kawa) and top-down (Benton) examinations of the direction of policy intervention, their work helps us understand how comparative policy analytical work can be undertaken vertically in both top-down and bottom-up trajectories.

Three contributions to studies of the regulation and governance of informal work – Nagle's *Picking Up*, Reno's *Waste Away*, and Millar's *Reclaiming the Discarded* – all focus on the governance of waste, but examine different work functions in different geographical regions. Nagle conducted an ethnography of sanitation workers in New York City (Nagle 2014), whereas Reno embedded himself in a US local landfill (Reno 2016), and Millar examined the lives of informal waste pickers in Rio de Janeiro, Brazil (Millar 2018). We could use my pull approach to peruse these authors' research to conduct cross-national (Reno and Miller) and cross-sectoral (Nagle and Reno) comparative waste governance analyses.

5. COMPARATIVE POLICY ANALYSIS AND POLICY STUDIES FROM A COMPARATIVE PERSPECTIVE: THE TWO FACES OF JANUS?

One of the important and most interesting issues that arise when we attempt to categorize the various ways in which ethnographic methods are applied to comparative policy analysis is the broad range of applications that various methods of comparison offer public administration, public policy, and public management. While I am wary of engaging in a categorization and

labeling war, from which we've seen very little productive discourse emerge, it is important to note that comparisons can be pursued in all areas of the policy sciences and public administration literatures. Comparative analysis can be applied to all these areas, and therefore we continue to see blurred definitions of what comparative public policy (Engeli, Rothmayr Allison, and Montpetit 2018), comparative policy studies (Engeli and Rothmayr Allison 2014), and comparative public administration (Van de Walle and Brans 2018) all mean. A recently published survey article marked the 20th anniversary of the *JCPA*, the premiere journal outlet for comparative work, by outlining the multiple meanings of comparative policy analysis (Van de Walle and Brans 2018).

Even though I am reluctant to conflate definitions presented by Van de Walle and Brans, Engeli, Rothmayr Allison, and Montpetit, and Radin and Weimer, after the careful display and thorough categorization their authors offer, I do argue that comparative policy analytical work can emerge from, and be undertaken through, systematic applications of various theoretical frameworks that inform all areas (public management, administration, policy) of the policy sciences. As Peters, Fontaine, and Mendez indicate, the purpose of comparative policy analysis is to help understand a broad range of phenomena, of which change is a fundamental one. I want to call special attention to the questions they raise about comparative analysis of policy change:

Change is a multidimensional problem which raises many questions for policy analysis. Why does it occur (or not)? Where does it come from? How can it be measured? When does it become irreversible? How can it be predicted? (Peters, Fontaine, and Mendez 2018: 135)

Tracing the causal mechanisms of policy change and following potential trajectories that can emerge from implementing a specific policy instrument can both be undertaken through the application of a comparative perspective at a micro-scale. Qualitative methods are particularly amenable to this undertaking. Of all these non-quantitative methodological tools, ethnography is the one that best provides insights about individual and group behavior that can then be used to inform policy design (Howlett and Mukherjee 2018).

One could also use process tracing in combination with ethnographic approaches to establish trajectories of causality and evaluating effects of instruments, programs and policies. These applications can be undertaken using a comparative perspective, very much along the lines of what the field of comparative politics does. In fact, one could easily argue that both fields (comparative policy analysis and comparative politics) are inextricably linked, as B. Guy Peters demonstrates. While these sub-fields seem to have followed diverging trajectories, Peters makes explicit the clear linkages that exist between the two scholarly camps (Peters 2018).

Researchers should be forewarned, of course, when attempting to use qualitative methods. There are several valid critiques that arise when engagement with the method and its application is done poorly. These criticisms usually focus on the possibilities of making mistakes by choosing a quantitative tradition and epistemology for a problem that requires a qualitative approach. The very nature of how we do research and the type of policy issue area we are trying to address have impacts on methodological choices we make. As Brower and colleagues indicate, researchers should be very careful when designing a study requiring a qualitative research strategy, primarily because the assumptions that quantitative approaches make do not apply to qualitative methods (Brower, Abolafia, and Carr 2000).

6. WHICH POLICY ISSUE AREAS USE ETHNOGRAPHY?

Perhaps the most well-known substantive policy area where ethnography is clearly and legitimately used to conduct and engage in comparative policy analysis is education. A quick Google Scholar search using the search terms “ethnography” and “policy analysis” immediately yields citations associated with educational policy on the first page of results. Certainly, ethnography is frequently applied in critical policy studies. While it is not a prerequisite that ethnographies have a critical component, the method emerges from a conversation between interpretivist and positivist methodological traditions in qualitative research.

Educational policy studies have used ethnographic approaches for a very long time now, in particular because of the critical nature of the method itself. Ethnography has a very critical approach to social investigations. Ethnographic research necessitates a clear acknowledgment of the researcher’s own positionality and a deep understanding of reflexivity as a mode of inquiry. Positionality and reflexivity are extremely important components (Katz 1994; Sultana 2007).

Ethnography is not the only method that critical education policy scholars use. In his analysis of policy historiography, policy archaeology, and policy genealogy, Gale outlines how different research methodologies contribute to our shared understanding of the factors that shape educational agendas, policy instruments, and outcomes. Gale’s analysis focuses on Australian higher education during the late 1990s (Gale 2001), but his insights can be extrapolated to other countries, issue areas, and temporal horizons. This demonstrates the value of this kind of work for comparative policy analysis. Gale’s typology of policy analytical methodologies helps us situate ethnography as a research method within a broad range of qualitative methods. While Gale’s paper isn’t about ethnography itself, his main comment does highlight the importance that the researcher’s own positionality has in undertaking comparative educational policy analysis. Making ethnographic research explicit about this issue is also fundamentally at the core of critical inquiry.

While there are some elements of positivism in ethnographic inquiry, the method itself and its theoretical grounding are clearly located closer to the critical side of the spectrum. Even though ethnography can be used to describe an observable phenomenon, it also can (and probably should) be used to reflect on practices of oppression and engage in critical examinations of power imbalances. As Levinson and collaborators outline, a critical approach to policy can lead to the democratization of policy processes as well (Levinson, Sutton, and Winstead 2009).

7. DOING ETHNOGRAPHY FOR COMPARATIVE POLICY ANALYSIS: AN APPLIED APPROACH

In this chapter, I pay special attention to the methodological strategy, and focus not only to the theoretical underpinnings of why we do ethnography or when we should use it, but also on how a policy analyst should enter the field and what tactics he/she should use. While the core of this chapter is centered on doing comparative policy analysis using ethnographic work, one could easily adapt these techniques to fit non-comparative, single-case study work. My main goal with this chapter is to help the reader navigate the broad range of strategies and techniques that ethnographers in various disciplines use and facilitate decision-making processes in which investigators must engage.

There are several choices that inherently accompany every research project. How to decide on the number of cases? How many individuals should engage in fieldwork and what tasks should be assigned? What are the ethics of engaging in studying vulnerable populations? Given that so many policy problems impact highly marginalized communities, how can a researcher minimize negative impacts on these communities? What kind of data should the researcher seek to find and extract from informants and contexts, and to what extent should he/she be engaging in action-research or participant observation that has a public/societal engagement component? In this chapter I address some of these questions, albeit I can't offer a more in-depth examination within the space allotted and scope of this volume.

Here I just would like to briefly outline strategies for each stage of the research cycle. While a detailed discussion of each one of these steps falls outside the scope of this chapter, I do want to ensure that the reader is aware of the broad range of activities that engaging in ethnographic work using a comparative policy analysis approach would entail. I list these stages below:

- a. positing a policy-analytical research question;
- b. choosing target populations/communities/countries/case studies;
- c. deciding on specific investigative approaches, fieldwork strategies, and team composition;
- d. preparing interview and open-ended question protocols;
- e. entering the field and building trust and rapport through engaging with the community;
- f. writing field notes that are specifically policy analytical;
- g. synthesizing data gleaned from field notes and analyzing these data with a comparative policy analytical lens; and
- h. reporting ethnographic data for policy makers.

I would also take the time to discuss the ethical, practical, pragmatic, and field-specific implications of using ethnographic approaches to comparative policy analysis. For this specific component, but also relevant to the others too, I focus on the three combined elements that make comparative policy analysis what it is: (a) *analysis* (that is, the systematic examination of information, data, and evidence to produce insights that can then explain phenomena); (b) *policy* (that is, an approach to solving public problems); and (c) *comparative* (therefore, implicitly drawing comparisons across observation, cases, approaches, units of analysis).

Ethnographic work facilitates analysis by providing rigorously collected empirical evidence. Data obtained through in-depth fieldwork can then be used to analyze individual policies' performances, their effects on specific target populations and potentially, evaluate the positive and negative consequences of policy implementation. For example, van Hulst has demonstrated that ethnography can be used to understand local governance by embedding oneself in cities' agencies and organizations (van Hulst 2008).

Van Hulst's examination of ethnographic research in local governments raises three very important issues regarding the epistemological and ontological contours of ethnography as a research method.⁴ The first one is discerning the research method's reliability and validity. How do we know that what the ethnographer has observed is reliable? Can we reproduce his/her observations and if so, what would be the way to do it? The second one is access to observed agents and actors. How do we guarantee that we will have access to those we need to observe in order to clearly and properly understand the inner workings of an agency within a particular context, as the local government example shows in the case of van Hulst? And the third one is the ontological and epistemological assumptions of ethnography as a research method. Often, ethnographic work is considered constructivist because of its inductive nature.

We normally do not establish hypotheses that can then be tested through ethnographic investigations of a specific phenomenon.

To address the first issues that I have indicated above, I bring back the Huey comparative policing study as well as Dahl's foundational text. While one could question that ethnography could have reliability and validity issues, I argue that undertaking systematic comparative analyses works instead to help reduce the problem of having ethnographic insights challenged. By using the same methodological approach (embedding oneself within each city for the same period, spending equal amounts of time and employing a systematic recording method for field notes for each one of the cases under study), researchers can work to assuage these issues. LeCompte and Goetz had raised this issue in the mid-1980s because of the increasing concern with lack of replicability, a feature of experimental work that helped improve how much we could trust results. We also need to discern the difference between external and internal validity. As these authors indicate:

external validity addresses the issue of whether independent researchers would discover the same phenomena or generate the same constructs in the same or similar settings. Internal reliability refers to the degree to which other researchers, given a set of previously generated constructs, would match them with data in the same way as did the original researcher. (LeCompte and Goetz 1982: 32)

Dahl's study of New Haven bureau chiefs demonstrates both internal and external validity. If a comparative policy analysis scholar were to undertake a similar study in a different jurisdiction, he/she could use the same methods as posited in the Dahl book and compare outcomes, in particular examining the hypotheses that Dahl used coming into the research.

To answer the second and third questions, I want to come back to what I believe makes ethnography a suitable method for comparative policy analysis. I argue that smaller-N, in-depth observational and participatory approaches to understanding policy processes across different cases have high explanatory power to discern potential causal pathways and developmental policy trajectories. Because ethnographers follow individuals and smaller groups throughout a longer/extended period, they are also able to trace how decision-making processes take place at a micro-scale. Larger-N methodologies are unable to capture individuals' reasoning processes, whereas continued contact with interviewees and key informants can provide more insight into how they made policy choices and/or brought issues to a specific forum. Take the issue of comparative policy agendas research. This area has seen enormous growth in recent years, and the Comparative Policy Agendas Project has facilitated large-scale, larger-N scale studies of how specific issues reach the agenda. How agendas are set can be studied with both quantitative and qualitative methodologies, but micro-level analyses can only be undertaken using a smaller-N research strategy. Given that ethnography necessitates but also facilitates in-depth connections between informants and the researcher, it is more likely that he/she will be able to trace decision-making processes throughout more easily than if we inferred them from a larger-scale study.

The Comparative Policy Agendas Project has a specific set of assumptions and hypotheses that can therefore be used by ethnographers to create research questions that they will be able to analyze as they engage with the field research. This is a different ontological and epistemological approach to what one traditionally would understand as ethnography. Instead of "letting the data speak" and inductively formulating propositions that could potentially then be testable through larger-N quantitative studies, we could set up a project where we come into the field with one specific question that we want to test, across two or three different jurisdictions, cities,

or communities. This is a positivist approach and one where using ethnography as a method can be justified because of the peculiar issue area that researchers are interested in studying. For example, comparative ethnographies of urban walking can provide us with insights into the patterns of social mobility within different demographics and geographical scales (Pierce and Lawhon 2015). We could set up specific case studies with great variation, testing how individuals walk within their neighborhoods in areas where there is a strong police force (perhaps making individuals feel safer) and in others where there is no policing (thus making pedestrians feel less safe). This is a positivist research design in the field of transportation and urban policy that can easily insert ethnographic work into the research methodology mix.

8. PROMISES AND PERILS OF THE APPLICATION OF ETHNOGRAPHY AS A RESEARCH METHOD FOR COMPARATIVE POLICY ANALYSIS

While in previous sections I have presented the virtues of using ethnographic methods in policy research through a comparative perspective, in this section I want to outline the potential this method has, but also the challenges that it poses for comparative policy analysts. Even though ethnography as a research strategy is highly involved and can enable researchers to create long-term bonds and establish relationships of trust between communities under study and scholars, it is also time-consuming and frequently expensive, not only in terms of financial investments but also researchers' personal time.

Ethnography as a research method has relatively solid popularity in sociology and human geography, but its applications in other social sciences and fields of study seem to be sparse, and sometimes questioned. Even though there are many excellent recent ethnographies that have proven to have societal relevance, wide global visibility, and the potential for important policy impact such as Kathy J. Cramer's political ethnography of rural voters in Wisconsin and Matt Desmond's devastating analysis of evictions in Milwaukee, some critics remain adamant about the negatives that accompany the use of ethnography in policy-relevant scholarship. This is partly in response to recent critiques of sociological work such as Alice Goffman's in-depth ethnography of over-policed African American communities in Philadelphia, partly a result of the overt trend of social sciences to focus on quantitative methods and large numbers of cases.

There are also valid critiques recently questioning researcher ethics in the use of this method, the (apparent) lack of reproducibility, and the boundaries and limits of what an ethnographer can capture within a certain temporal and spatial horizon. Many of these critiques are quite valid regardless of whether the method is used in anthropological studies, sociological analyses, or public administration scholarship. In this section, I review recent critiques of ethnography as a research method, while maintaining the position that there is still much potential for this methodological approach for the comparative study of policy development, and for conducting comparative policy analysis.

There are critiques of ethnography that center around generalizability and the validity of broader generalizations drawn from a primarily inductive approach. Given the extraordinarily low number of cases that can be analyzed through ethnographic approaches if there is only one researcher, critics of the method argue that there is not a lot of room for generalization from the specificity that ethnography brings along.

Other criticisms focus on *external validity*, *reliability*, and *replicability*. These critiques center around the question of whether one can draw causal claims from ethnographic observation, given the deeply personal nature of this method and the implicit but clear potential for bias. Ethnographers bring their own experiences to the field and can potentially also show biases and maintain priors instead of looking for puzzles, uncovered patterns, and paradoxes. I have explicitly specified that there is a potential and therefore a possibility for a comparative policy analyst using ethnographic approaches to show bias and to have that bias applied to his/her understanding of a phenomenon. But there is conflicting evidence on whether this is a frequent occurrence. If we seek to demonstrate that researchers and analysts have inherent implicit (or even explicit) biases that they can then apply to their case study, we will need to be very specific in pointing out the mechanisms through which biases are transmitted and affect the validity of ethnographic research.

One of the major critiques of ethnography has come from an ethnographer himself. Rist's concern is that, because of "the low price of admission", anybody who declares themselves to be an ethnographer can be perceived as such, even if they don't actually have any training or experience (Rist 1980). This worry has been echoed elsewhere, particularly because conducting an ethnography "appears easy". It would seem as though simply by embedding oneself without any systematic method or robust technique or strategy for rigorous data collection, one would be able to perform ethnographic work. Nothing is further from the truth. Simply parachuting into a community and spending a few days/weeks/months talking to residents does not entail ethnographic work and cannot be considered as such. Hammersley has called attention to this problem by indicating that disagreement about what ethnography is and how it should be used could potentially lead to even more attacks on the method (Hammersley 2018) and less interest not only from sociologists, political scientists, and human geographers, but from other disciplines as well. I agree with this take only partially because critiques are necessary for survival and growth of a research method.

I recognize three main concerns with ethnography that can be of relevance to comparative policy analysis. As far as I can tell, researchers from other disciplines and ontological traditions are concerned with reproducibility, transparency, reliability, validity, and ethics. The first concern I'd like to tackle in this discussion is the combination of issues of reliability and validity. How do we decide what kind of ethnographic data is valid and whether the ethnographer is reliable? What kind of data collection, systematization, and storage practices are necessary to make ethnographic reports valid? These are valid questions and remain important if we are to conduct comparative analyses of policy issues that rely on the perceptions of a few researchers. One potential mechanism to alleviate this concern would be to undertake intercoder reliability measurements. This strategy works at the analysis stage, when themes and codes have already been generated (Ryan and Bernard 2003). However, this process probably wouldn't work as easily or as well when the number of researchers is high or when the ethnographies have been multi-sited and different researchers have conducted components of the same project.

On the second issue, it is clearly important to have a more reliable science. Reproducibility and replicability both ensure that our research process can be reproduced, and the data generation process replicated, leading to more robust conclusions about the mechanisms at play. With ethnography, this replicability becomes a tad problematic because no two ethnographers are the same. Individual positionality and reflexivity are distinct across scholars and analysts, and therefore interpretation of the same event and experience can vary. This is a feature of the

method, and not a bug. Fieldworkers could potentially store and share their field notes, and ask other researchers to interpret from recorded interview transcripts, but it is quite likely that these investigators will reach different conclusions. Moreover, fostering transparency and data sharing also improves reproducibility. Obvious concerns such as protecting vulnerable populations and ensuring data privacy need to be addressed, of course, and are not minor items on the agenda, specifically because policy analytical work usually deals with sensitive information about target populations, stakeholders, budgets, and policy actions. Thus, researchers will need to ponder how to balance transparency with data protection.

On the ethical conundrum that ethnography poses, there are at least two elements that intersect previous concerns. The first one is whether it is ethical to destroy field notes when marginalized populations' privacy is at risk. This worry first arose as a result of numerous critiques of Alice Goffman's award-winning *On the Run*, which is the result of her six-year-long ethnography of over-policed African American communities in Philadelphia, Pennsylvania, in the United States. While lauded in many forums, and winning Best Dissertation Prize in the American Sociological Association's annual meeting, Goffman was also heavily criticized because she disposed of her field notes to avoid having them subpoenaed. These critiques are quite valid because there is no way to verify any of her claims, given the strong degree of anonymization that her data underwent. Goffman's tribulations became a warning sign that other ethnographers heeded. Matthew Desmond, in his ethnography of evicted individuals and families in Milwaukee, had an external verification process throughout the process of generating his award-winning *Evicted*.

The second ethical issue concerns ethnographers' treatment of vulnerable communities and individuals. How to engage in fieldwork in areas where extreme deprivation and abject poverty are rampant? What kind of strategies should researchers use to prevent harm and at the same time, ensure visibility and avoid erasure? Comparative policy analysis has enormous potential for positive societal impact by conducting what Pacheco-Vega and Parizeau call attention to, doubly-engaged social science. This is a term coined by Theda Skocpol, which these authors apply to suggest a doubly-engaged ethnographic approach (Pacheco-Vega and Parizeau 2018).

I should note, however, that more recent, innovative approaches to ethnography, specifically quantitative ethnography and computational ethnography have been praised as having potential to address many of the shortcomings of ethnographic techniques, and to respond to criticisms that this method faces (Trondman 2000). Moreover, a more pragmatic approach to causality in ethnography can potentially strengthen perceptions of validity of inferences drawn from in-depth observational studies through fieldwork. This pragmatic approach is applied by Tavory and Timmermans in their study of reactions of parents and clinicians to positive newborn screening results (Tavory and Timmermans 2013).

9. CONCLUSION: EMBRACING ETHNOGRAPHY IN COMPARATIVE POLICY ANALYSIS?

In this chapter, I have outlined the many ways in which ethnography can be used in comparative policy analysis. I have presented not only a basic overview of the method, but I have also examined whether choosing a small-N qualitative methodology such as ethnography is beneficial or detrimental to comparative studies of public policies. One clear and obvious advantage of using ethnography in comparative policy analysis is the benefit of drawing insights from

marginalized and vulnerable communities which would otherwise be rendered invisible. In-depth immersion enables the researcher to absorb key details from target populations that would not be easy to determine through other methodologies.

When should we use ethnography in comparative policy analysis? Clearly, as the rest of the chapters presented in this volume show, and as I argue in this chapter, these in-depth, detailed insights cannot be drawn from large-N studies, but even in smaller-N analyses, ethnography is clearly superior to other qualitative methods as it enables the researcher to narrow the focus in such a way that the lived experiences of individual participants can be better, more accurately and clearly reported. From a policy analytical perspective, ethnography facilitates robust policy design and effective instrument choice by connecting target populations with policy analysts and decision makers.

Who should be responsible for conducting ethnographic comparative policy analysis? While policy analysts can be ethnographers themselves who could conduct field visits and in-depth observational tasks as well as analyze and synthesize data, some projects may specify that trained fieldworkers be the ones conducting ethnographic work. While governments may conduct these studies on their own, it is also likely that they will hire someone, such as a private firm, a university group, or a subcontractor. In any case, when conducting ethnographic work for comparative policy analysis, it is imperative that everyone working in the field undertake work that is ethically robust, methodologically sound, and substantively insightful.

Why should ethnography be used in comparative policy analysis when other qualitative techniques are also available to researchers and could potentially yield similar insights to those drawn from in-depth observational fieldwork? As I outline in the chapter, ethnography allows researchers to be physically and intellectually closer to their subject of study. While structured and semi-structured interviews can help draw relevant insights from key informants, the embedded nature of ethnographic work facilitates repeated interaction with participants, and enables the researcher to capture observational data throughout a longer period of time.

How should we implement ethnography in comparative policy analysis? I attempted to answer this question throughout the chapter, and it would be glib to attempt to provide a short, witty answer here. But one of the most important elements to consider is applying a framework that facilitates community engagement with policy analysts, bureaucrats, and politicians. Such a model of ethnography is championed by Pacheco-Vega and Parizeau (2018), where not only are vulnerable and marginalized individuals and groups protected, but also policy issues that are relevant to them can be brought to light and amplified (Pacheco-Vega and Parizeau 2018). Their “doubly-engaged ethnography” framework can provide a template for comparative policy analysis that is not only theoretically insightful and empirically robust but also socially responsible.

NOTES

1. Though I note that within-case variation usually is examined using process tracing.
2. For the reason I noted earlier in the chapter, I looked at how ethnography was used within the policy sciences. Therefore, in this section I do not make a distinction between public administration, public policy, and public management as my intent here is to demonstrate that the method is being applied. In later sections of the chapter I explain how ethnographic methods are applied to specific policy issue areas including, but not limited to education.
3. I am grateful to B. Guy Peters for pointing me to this specific work.

4. I am very grateful to B. Guy Peters and Guillaume Fontaine for raising these three issues. Combined, they provide the reader with a much clearer picture of how ethnography can be applied from a clearly positivist vantage point.

REFERENCES

- Anand, N. (2017). *Hydraulic City: Water & the Infrastructures of Citizenship in Mumbai*. Durham, NC and London: Duke University Press.
- Becker, H. S., Gans, H. J., Newman, K. S., and Vaughan, D. (2004). On the value of ethnography: sociology and public policy: a dialogue. *Annals of the American Academy of Political and Social Science* 595: 264–76.
- Benton, A. (2015). *HIV Exceptionalism: Development through Disease in Sierra Leone*. Minneapolis, MN: University of Minnesota Press.
- Brower, R. S., Abolafia, M. Y., and Carr, J. B. (2000). On improving qualitative methods in public administration research. *Administration & Society* 32 (4): 363–97.
- Burt, B. (1998). What is ethnography? *Journal of Museum Ethnography* 10: 10–14.
- Cantú, F. (2018). *The Line Becomes a River: Dispatches from the Border*. New York: Riverhead Books.
- Cappellaro, G. (2017). Ethnography in public management research: a systematic review and future directions. *International Public Management Journal* 20 (1): 14–48.
- Clifford, J. and Marcus, G. E. (1986). *Writing Culture: The Poetics and Politics of Ethnography*. Berkeley: University of California Press.
- Crossa, V. (2012). Disruption, yet community reconstitution: subverting the privatization of Latin American plazas. *GeoJournal* 77 (2): 167–83.
- Dahl, R. (2005). *Who Governs? Democracy and Power in an American City* (2nd edition). New Haven, CT: Yale University Press.
- Davies, J. S. (2002). The governance of urban regeneration: a critique of the “governing without government” thesis. *Public Administration* 80 (2): 301–22.
- De Leon, J. (2015). *The Land of Open Graves: Living and Dying on the Migrant Trail*. Oakland, CA: University of California Press.
- Desmond, M. (2016). *Evicted: Poverty and Profit in the American City*. New York: Broadway Books.
- Dubois, V. (2009). Towards a critical policy ethnography: lessons from fieldwork on welfare control in France. *Critical Policy Studies* 3 (2): 221–39.
- Engeli, I. and Rothmayr Allison, C. (Eds.) (2014). *Comparative Policy Studies: Conceptual and Methodological Challenges*. Basingstoke: Palgrave Macmillan.
- Engeli, I., Rothmayr Allison, C., and Montpetit, É. (2018). Beyond the usual suspects: new research themes in comparative public policy. *Journal of Comparative Policy Analysis: Research and Practice* 20 (1): 114–32.
- Erasmus, E. (2014). The use of street-level bureaucracy theory in health policy analysis in low- and middle-income countries: a meta-ethnographic synthesis. *Health Policy and Planning* 29 (Suppl. 3): iii70–iii78.
- Gale, T. (2001). Critical policy sociology: historiography, archaeology and genealogy as methods of policy analysis. *Journal of Education Policy* 16 (5): 379–93.
- Geva-May, I., Hoffman, D. C., and Muhleisen, J. (2018). Twenty years of comparative policy analysis: a survey of the field and a discussion of topics and methods. *Journal of Comparative Policy Analysis: Research and Practice* 20 (1): 18–35.
- Goffman, A. (2014). *On the Run: Fugitive Life in an American City*. Chicago, IL: University of Chicago Press.
- Guarneros-Meza, V. (2009). Mexican urban governance: how old and new institutions coexist and interact. *International Journal of Urban and Regional Research* 33 (2): 463–82.
- Hackenberg, R. A. (1985). Bringing theory back in: steps toward a policy science of applied anthropology. *American Behavioral Scientist* 29 (2): 205–28.
- Hammersley, M. (2018). What is ethnography? Can it survive? Should it? *Ethnography and Education* 13 (1): 1–17.

- Hoover, E. (2017). *The River Is in Us: Fighting Toxics in a Mohawk Community*. Minneapolis, MN: University of Minnesota Press.
- Howlett, M. and Mukherjee, I. (2018). The contribution of comparative policy analysis to policy design: articulating principles of effectiveness and clarifying design spaces. *Journal of Comparative Policy Analysis: Research and Practice* 20 (1): 72–87.
- Huby, G., Harries, J., and Grant, S. (2011). Contributions of ethnography to the study of public services management. *Public Management Review* 13 (2): 209–25.
- Huey, L. (2007). *Negotiating Demands: The Politics of Skid Row Policing in Edinburgh, San Francisco, and Vancouver*. Toronto: University of Toronto Press.
- Johnson, D. C. (2009). Ethnography of language policy. *Language Policy* 8 (2): 139–59.
- Katz, C. (1994). Playing the field: questions of fieldwork in geography. *The Professional Geographer* 46 (1): 67–72.
- Kawa, N. C. (2016). *Amazonia in the Anthropocene: People, Soils, Plants, Forests*. Austin, TX: University of Texas Press.
- Kuus, M. (2013). Foreign policy and ethnography: a sceptical intervention. *Geopolitics* 18 (1): 115–31.
- Ladner, S. (2014). *Practical Ethnography*. Walnut Creek, CA: Left Coast Press.
- LeCompte, M. D. and Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of Educational Research* 52 (1): 31–60.
- Levinson, B. A. U., Sutton, M., and Winstead, T. (2009). Education policy as a practice of power: theoretical tools, ethnographic methods, democratic options. *Educational Policy* 23 (6): 767–95.
- Lora-Wainwright, A. (2017). *Resigned Activism: Living with Pollution in Rural China*. Cambridge, MA: MIT Press.
- Maginn, P. J. (2006). Urban policy analysis through a qualitative lens: overview to special issue. *Urban Policy and Research* 24 (1): 1–15.
- Mah, A. (2010). Memory, uncertainty and industrial ruination: Walker Riverside, Newcastle upon Tyne. *International Journal of Urban and Regional Research* 34 (2): 398–413.
- McGranahan, C. (2014). What is ethnography? Teaching ethnographic sensibilities without fieldwork. *Teaching Anthropology* 4: 23–36.
- Mears, A. (2011). *Pricing Beauty: The Making of a Fashion Model*. Berkeley, CA: University of California Press.
- Millar, K. M. (2018). *Reclaiming the Discarded: Life and Labor on Rio's Garbage Dump*. Durham, NC and London: Duke University Press.
- Mosley, L. (Ed.) (2013). *Interview Research in Political Science*. Ithaca, NY: Cornell University Press.
- Nagle, R. (2014). *Picking Up: On the Streets and Behind the Trucks with the Sanitation Workers of New York City*. New York: Farrar, Straus & Giroux.
- Okongwu, A. F. and Mencher, J. P. (2000). The anthropology of public policy: shifting terrains. *Annual Review of Anthropology* 29: 107–24.
- Pacheco-Vega, R. and Parizeau, K. (2018). Doubly-engaged ethnography: opportunities and challenges when working with vulnerable communities. *International Journal of Qualitative Methods* 17 (1): 1–13.
- Pachirat, T. (2013). *Every Twelve Seconds: Industrialized Slaughter and the Politics of Sight*. New Haven, CT: Yale University Press.
- Paluck, E. L. (2010). The promising integration of qualitative methods and field experiments. *Annals of the American Academy of Political and Social Science* 628 (1): 59–71.
- Peters, B. G. (2018). Comparative politics and comparative policy studies: making the linkage. *Journal of Comparative Policy Analysis: Research and Practice* 20 (1): 88–100.
- Peters, B. G., Fontaine, G., and Mendez, J.-L. (2018). Substance and methods in the comparative study of policy change. *Journal of Comparative Policy Analysis: Research and Practice* 20 (2): 133–41.
- Pierce, J. and Lawhon, M. (2015). Walking as method: toward methodological forthrightness and comparability in urban geographical research. *The Professional Geographer* 67 (4): 1–8.
- Ranganathan, M. (2014). “Mafias” in the waterscape: urban informality and everyday public authority in Bangalore. *Water Alternatives* 7 (1): 89–105.
- Ranganathan, M. (2018). Rethinking urban water (in)formality. In K. Conca and E. Weinthal (Eds.), *The Oxford Handbook of Water Politics and Policy*. Oxford: Oxford University Press, 310–24.
- Reno, J. O. (2016). *Waste Away: Working and Living with a North American Landfill*. Oakland, CA: University of California Press.

- Rist, R. C. (1980). Blitzkrieg ethnography: on the transformation of a method into a movement. *Educational Researcher* 9 (2): 8–10.
- Rist, R. C. (1981). On the utility of ethnographic research for the policy process. *Urban Education* 15 (4): 485–94.
- Ryan, G. W. and Bernard, H. R. (2003). Techniques to identify themes. *Field Methods* 15 (1): 85–109.
- Stuart, F. D. (2016). *Down, Out, and Under Arrest: Policing and Everyday Life in Skid Row*. Chicago, IL: University of Chicago Press.
- Sultana, F. (2007). Reflexivity, positionality and participatory ethics: negotiating fieldwork dilemmas in international research. *ACME* 6 (3): 374–85.
- Tavory, I. and Timmermans, S. (2013). A pragmatist approach to causality in ethnography. *American Journal of Sociology* 119 (3): 682–714.
- Trondman, M. (2000). Manifesto for ethnography. *Ethnography* 1 (1): 5–16.
- Van de Walle, S. and Brans, M. (2018). Where comparative public administration and comparative policy studies meet. *Journal of Comparative Policy Analysis: Research and Practice* 20 (1): 101–13.
- van Hulst, M. J. (2008). Quite an experience: using ethnography to study local governance. *Critical Policy Studies* 2 (2): 143–59.
- Vogtz, F. (2002). No ethnography without comparison: the methodological significance of comparison in ethnographic research. In G. Walford (Ed.), *Debates and Developments in Ethnographic Methodology, Volume 6*. London: Elsevier Science, 25–32.
- Walt, G., Shiffman, J., Schneider, H., Murray, S. F., Brugha, R., and Gilson, L. (2008). “Doing” health policy analysis: methodological and conceptual reflections and challenges. *Health Policy and Planning* 23 (5): 308–17.
- Wedel, J. and Feldman, G. (2005). Why an anthropology of public policy? *Anthropology Today* 21 (1): 1–2.
- Wedel, J. R., Shore, C., Feldman, G., and Lathrop, S. (2005). Toward an anthropology of public policy. *Annals of the American Association of Political and Social Science* 600 (1): 30–51.
- Wolf, A. and Baehler, K. J. (2018). Learning transferable lessons from single cases in comparative policy analysis. *Journal of Comparative Policy Analysis: Research and Practice* 20 (4): 420–434.
- Wolf, S. (2017). *Mano Dura: The Politics of Gang Control in El Salvador*. Austin, TX: University of Texas Press.
- Wylie, S. A. (2018). *Fractivism: Corporate Bodies and Chemical Bonds*. Durham, NC: Duke University Press.
- Yanow, D. (2000). *Conducting Interpretive Policy Analysis*. Thousand Oaks, CA: Sage.

19. Using Q methodology in comparative policy analysis

Astrid Molenveld

1. INTRODUCTION

Numerous scholars and practitioners claim that objective policy measures free from preferences do not exist (Brown 1974; Howlett 2009; Peters and Hoornbeek 2005; Scharpf 2000; Wlezien 2004). These authors contend that ideas about the nature of knowledge, envisioned outcomes of policy, spending and politics associated with policy solutions differ between stakeholders, politicians, civil servants, and citizens (Peters 1998). This argument emphasizes that, both societal problems and their solutions are often viewed by multiple stakeholders from different angles because of their particular interests, values, and worldviews. This is typically how wicked and complex policy problems are described (Rittel and Webber 1973; Termeer, Dewulf, and Breeman 2013).

The basis of these differences is fundamental – people view and perceive complex policy issues profoundly differently, and therefore policy solutions are inherently value-laden and biased by different problem frames (Brown 1974; Rittel and Webber 1973). This becomes clear when bringing stakeholders and public servants with different fields of expertise and knowledge together in debate about possible policy instruments and solutions (Ansell and Gash 2007). A policy solution that satisfies one organization might be objectionable, or create a problem for another individual or organization. These objections are difficult to comprehend, but important to unravel, especially as personal perspectives form “perceptual lenses” (Allison 1971) through which ideas, frames, or policy stories are created around specific issues. As a consequence of the above, many policy scientists are interested in *what* frames can be identified and *how* they influence debate, collaboration, and public policy decision making. In addition, in public administration practice, an increase in participation of non-governmental stakeholders in policy decision-making processes can be observed. This new *modus operandi* compels the public sector to be inclusive and open to an array of stakeholders with diverse opinions, to take into account minority voices, and to give stakeholders the opportunity to be involved in policy debates.

The two fundamental issues outlined above, i.e. *policy preferences* and *increased participation*, may explain why Q methodology, which makes subjectivity more explicit, is gaining popularity in the field of comparative policy analysis. Q methodology was developed by William Stephenson to study subjectivity in 1935. Today, the methodology has gradually (see Appendix) found its place in the toolbox of policy scientists, and it is especially useful for conducting comparative analyses. The method is comparative by nature as it seeks to uncover significantly different patterns in people’s viewpoints (Van Exel and De Graaf 2005; Watts and Stenner 2012; Wolsink 2004: 2682). A Q-methodological study often begins with exploring and developing insights into the multiple viewpoints present in a particular population about a certain topic. Based on the viewpoints within this population, the researcher develops

knowledge about the breadth of the debate, typically called the *concourse*. The researcher attempts to describe the concourse with a representative set of statements (which is a subset of the entire debate – Webler, Tuler, and Krueger 2001) which the same, or different respondents are asked to sort on a scale in a later stage of the research.

Respondents express their opinion about the issue debated, through sorting statements or words (the Q-sorting process) on their individual opinion. Rarer, especially in policy analysis literature, are studies based on photos, pictograms, or maps. Afterwards, the researcher looks for commonalities (by correlating the individual sorts). The researcher can, based on where respondents place particular statements or photos, identify patterns. Assisted by a factor analysis, the researcher must interpret subsequently the patterns of argumentation revealed by the respondents who sorted the statements.

Q methodology consists of approximately six steps, namely:

- Gaining insight into the concourse
- Representing the debate: Q-sample
- Selection of participants: P-sample
- Ranking by participants
- Analysis: defining the discourses
- Drawing conclusions based on these discourses.

This chapter explains the method and a range of applications of Q methodology through these steps. Although the approaches, goals, and research questions in Q-methodological applications differ, they also show clear similarities. Thus, in this chapter, the method *state of play* in comparative policy analysis is clarified. On the basis of a review of 40 articles and the most prominent Q-methodology textbooks and writings, this chapter distinguishes different elements in which Q-methodology applications in comparative policy research vary. This chapter illustrates different conceptual ideas, approaches, and lessons learned from these applications, which may be of interest to researchers planning a Q-methodological study.

2. PHILOSOPHY

Q methodology is a methodology designed to study the perceptions and prevalent discourses that exist within a given population (Brown 1993; Van Exel and De Graaf 2005), e.g. a policy community, a governance network trying to govern a common pool resource, or civil servants in the center of government. Its specific ontological stance, philosophy, and method, make it quite distinct from other methodologies (Ramlo and Newman 2011, 2015; Stenner, Watts, and Worrell 2007). It has many qualitative aspects and uses statistical analysis to uncover individuals' preferences (Brown 2008). Therefore, according to some, Q methodology has a different ontological basis (Ramlo and Newman 2015) compared to other methods. Susan Ramlo and Isadore Newman (2015) explain that some scholars call Q methodology a constructivist (or a qualitative) method, while some call it a positivist (or a quantitative) method. Moreover, they note that Stephenson himself stated that it is meant to study subjectivity in an objective way (Ramlo and Newman 2015: 174). Q methodology is, in essence, a method that assesses whether different majority and minority discourses exist within a community (Stenner, Watts, and Worrell 2007) using quantitative factor analysis to indicate possible areas of (dis)agreement. Because this methodology involves studying discourse, which is qualitative, through the

use of a systematic (factor) analysis, which is quantitative, it could be called a mixed-method. This systematic process of assessing qualitative data in a quantitative way may lead some researchers to believe that Q methodology is a positivistic approach, but this methodology does have a strong interpretative nature as well. This becomes clearer once two important steps of the methodology are explained: how the concourse is built *ex ante* and how factors are interpreted *ex post*.

A more thorough discussion on the method used in Q methodology is provided in the following section and examples of applications in comparative public policy are used to illustrate the method's steps. These applications were selected based on a systematic search for scholarly literature applying Q methodology. The execution of this selection process is explained in the Appendix.

3. HOW TO EXECUTE AND ANALYZE A Q-METHODOLOGY STUDY

Stephenson (1902–89) was a psychologist and used his Q-technique (published in 1935 in *Nature*) in clinical settings to assess individuals (Van Exel and De Graaf 2005). He was a student of Charles Spearman and was convinced that individuals should be studied holistically, not through analysis of individual items or personal characteristics (Watts and Stenner 2012). Therefore, he advocated and developed a method for the systematic study of subjectivity (Stenner, Watts, and Worrell 2007: 212). Where Spearman developed his R-correlation, which is an *item* analysis, Stephenson developed an inverted analysis, which enabled him to compare individuals and their viewpoints (Watts and Stenner 2012) as if they were “the variables”. This *person* analysis reveals which respondents express similar viewpoints and have thus similar positions about a topic of debate (Van Exel and De Graaf 2005). This mode of analysis, leading to unraveling significant different discourses, makes Q methodology an attractive technique for policy analysts interested in determining why conflicts emerge, which frames of reference exist, and how policy specificities are perceived.

Many of the applications of Q methodology are geared towards large groups of respondents, and this is particularly true in public policy analysis. However, when Q methodology was first applied it was used to explore the perspective of an individual on multiple issues (Brown 1993; Stenner, Watts, and Worrell 2007). Thus, Q methodology can be applied to conduct:

- *Intensive individual studies*: to reveal how an individual thinks about different constructs. In this type of study the researcher can use the same set of statements but vary their object, e.g. a civil servant is asked to sort the same set of statements about how he or she sees various users of a particular public service, e.g. a low-income single mother or a middle-aged migrant man.
- *Population studies*, which focus on a larger group and study significant differences between respondents. All articles under review in this chapter apply this approach, and the remainder of this chapter therefore relates to population studies.

Without going into excessive detail, a short overview of the methodology is provided below. However, the main focus here is on how the method is applied and its potential for comparative policy analysis. More information on the method can be found in the excellent work of McKeown and Thomas (2013), Brown (1993), Watts and Stenner (2005; 2012), Ramlo and

Newman (2011), Jeffares and Skelcher (2011), and Van Exel and De Graaf (2005), as well as many others.

3.1 Gaining Insight into the Concourse

The first step in Q methodology is defining the *concourse*, i.e. the debate about a certain issue. In popular speech, the word *concourse* means something akin to “the crossroads” or “the place where different streets meet” (Brown 1980; Watts and Stenner 2012). The Latin word *concurus* refers to entities running together (Davies, Blackstock, and Rauschmayer 2005). Both the idea of the crossroads and the notion of “a place” are important for understanding Q methodology. A debate about a particular issue often consists of multiple patterns of thought (cf. the crossroads), which meet and converge on different points (cf. places).

In debates that are highly charged – such as climate change policy – the different sides of the debate are often clear, which leads to ideas on the part of the researcher about what particular patterns might look like. Other debates might be less well known and more opaque for the researcher. This could include, for instance, viewpoints on poverty policy or windmills. The insight that Q methodology can offer in debates on particular policy plans, programs, or communities makes it useful for policy analysis (Dickinson et al. 2014: 837).

Scholars apply different approaches to gain insight into the concourse. Regarding the sources used to define the concourse, researchers often turn to existing concepts from literature, policy documents, interviews, focus groups, mass media, “gray literature”, etc. Studying the population a priori ensures that the statements shown to the respondents are *representative* for the concourse or discourses that exist within the population. In Q methodology, *representative* refers to the representativeness of the statements for the concourse, not the representativeness of the participants for the population. However, while the statements must *represent* the debate, they are not exhaustive and rather are a subset of the discourse as a whole (Webler, Tuler, and Krueger 2001). Understanding the true debate is of key importance, because the factors identified in a later stage are based on the ordering of the statements. If the statements do not represent both conflicting and overlapping areas of thought, the analysis and factors will be less clear and less prominent. However, most important is a true understanding of which statements are *salient* for the participants rather than bias them towards the categories of interest of the researcher.

To define the true debate, most scholars do sketch the dominant ideas that reappear in scholarly literature. Some researchers take these ideas as a basis for defining the breadth of the concourse. In this regard, one can clearly distinguish two types of approaches in comparative public policy in understanding the concourse: inductive and deductive. In the 40 articles studied for this chapter, scholars use Q methodology to further theory on a particular topic (*inductive*), as well as to confirm theory or theoretical concepts (*deductive*). When applying Stephenson’s ideas rigorously, the researcher has to base his or her statements on documents, articles, interviews, focus groups, newspapers, or any account of the possible concourse and define it *inductively*, instead of predefining categories (Curry, Barry, and McClenaghan 2013; Dimitrova and Kortenska 2017; Ockwell 2008). After this analysis the researcher identifies the existing viewpoints and attempts to understand them, rather than confirm overlap between preexisting concepts and ideas. Some researchers go even further, stating that Q methodology is designed to *avoid ex ante* description of theoretical concepts and “psychometric” tests

(Dickinson et al. 2014; Watts and Stenner 2005). Typically, Q researchers do not formulate hypotheses for the reasons described above.

Table 19.1 Differences between interpretations and Q methodology

	Qualitative interpretation	Q methodology	Quantitative interpretation	Author
Starting point	Theoretical or case-based expectations	Discourse or concourse in a population or community	Preconceived hypotheses	(Curry et al. 2013; Ockwell 2008)
Outcome/output	Contextualized in-depth knowledge	Subjective viewpoints	Relationships between variables	(Watts & Stenner 2005)
Role of the researcher/participant	Both, in close contact	Participant led	Researcher led, at a distance	(Curry et al. 2013; Ockwell 2008)

The aforementioned points illustrate the difference in viewpoints regarding what knowledge is and how to conduct research. This is essentially the difference between constructivism and positivism, but it might well be the difference between theory-building and theory-testing. Instead of wide generalization, Q methodology is geared especially towards theory-building for policy design, and thus conceptual generalization (Watts and Stenner 2012). In Table 19.1 the distinct features of Q methodology are presented. This is only one way to illustrate the differences and there may be other ways to distinguish between various methods.

3.2 Representing the Debate: Q-Sample

As the researcher progresses through his or her study, the information obtained about the debate (i.e. concourse), must be refined by the researcher into *statements*. Sorting statements is the manner in which participants who take part in the study express their thoughts about the topic. These statements should represent opinions, not facts, as the method tries to reveal discourse and subjectivity. The set of statements – called the Q-sample – is prominent and important, as it provides the basis for the discourses found in the subsequent analysis.

Most policy analysts attempt to identify statements from the discourse, with different methods, to ensure that they select statements that are *salient* to the respondents. These methods include pilot studies (e.g. Cuppen 2012; Curry, Barry, and McClenaghan 2013; Nekola 2012; Pelletier et al. 2000), a Delphi study (Nikraftar 2013), coding with multiple researchers, asking practitioners for feedback, etc. Two approaches in the review go even further to make certain they understand the debate and account for its most salient topics. First, Peritore and Peritore (1990) *include a researcher* to sort the statements in order to review and understand the statement's position and discourse. Second, de Wijs, Witte, and Geertman (2016) include *two empty cards* to give respondents the opportunity to include their own salient points if they feel that the researcher has omitted something highly relevant. This latter approach, however, may lead to a situation where it is no longer possible to conduct factor analysis.

In order to create a systematic selection of statements, most researchers use some type of thematic clustering (or sometimes even a matrix). One of the more well-cited Q-articles in political science is that of Dryzek and Berejikian (1993). They cluster *inductively gathered* statements in a *deductively informed* matrix, which consists of *substance* elements (rows) and *types of arguments* (columns). To determine the final set of statements, single statements are

placed within each cell. This aligns with what McKeown and Thomas (2013) call “a factorial design”, which is de facto a list with thematic clusters that illustrates how statements are derived. Although most researchers account for the number and salience of statements, Van Exel and De Graaf (2005: 5) emphasize that the selection of statements is partially based on the impression of the researcher (see also Brown 1980).

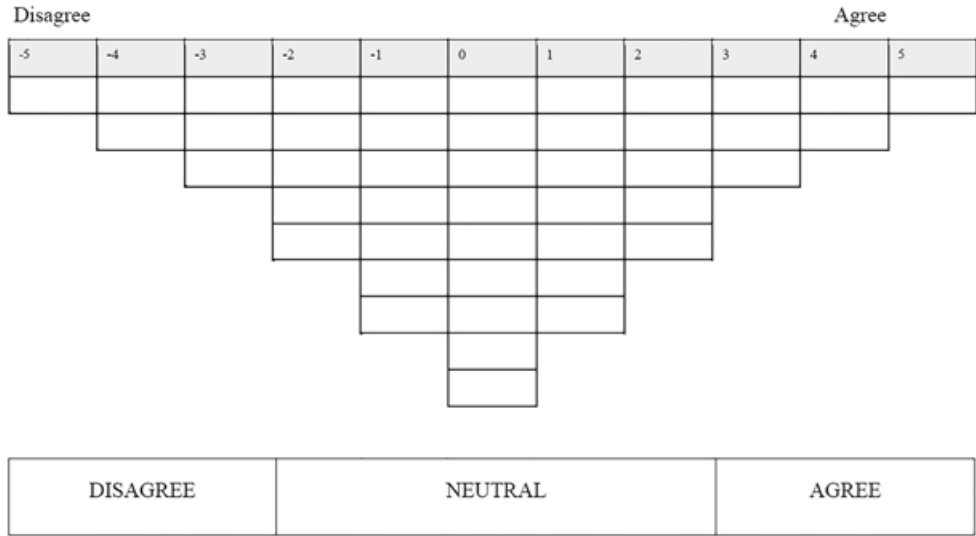


Figure 19.1 Example of forced normal distribution

On average, researchers select 25–50 statements from the concourse. In the studies on comparative policy the minimum number of statements is 24 (Nijnik, Nijnik, and Bizikova 2009) and the maximum is 80 (Eeten 2001). While in R-oriented methods researchers have to explain in depth how many respondents they included and what kind of sample they represent, Q methodologists must account for their selection of statements. Once the statements are chosen, the researcher determines the number of statements that could fit a normal (or quasi-normal) distribution (e.g. as in Figure 19.1). Although all articles included in this chapter use a *forced distribution*, this is not necessary. Brown (1980: 288–9) even illustrates that the effect of forced distributions can be neglected. Nevertheless, forced distribution can make it easier for a respondent to choose extremes and subsequently fill the grid with more neutral statements rather than sorting the entire set in a rank order that is not forced (Watts and Stenner 2012). When designing the distribution, a researcher does not need to be concerned with how many statements are under each point of the scale too much (i.e. the steepness). Regarding the steepness of the normal (or quasi-normal) distribution, there is generally one rule that researchers follow, namely if the respondents are very knowledgeable the researcher can make the distribution more flat (Watts and Stenner 2012). A steeper distribution allows respondents to place more statements in the middle rows, which represent the statements on which he or she is neutral. The most frequently used software, PQmethod (developed by Peter Schmolck),¹

which is specifically geared towards analysis of “forced” distributions, necessitates this distribution. Another package, Ken-Q (a desktop web application developed by Shawn Banasick²), is more flexible and also allows unforced distributions. Both packages are offered for free, online.

3.3 Selection of Participants: The P-Sample

Often, *comparison* is the reason policy analysis researchers engage in a Q-methodology study. This is done in order to discover (and to understand) the multifaceted and conflicting concepts that may exist among a group of participants. Frequently, these studies describe wicked issues in which many stakeholders have conflicting interests, problem frames, and strategies. The majority of articles study governance networks or deliberative fora, with a focus on decision making in the energy, water, or environmental sector. More than 75% of the articles address conflicts or difficulties in understanding others’ frames of reference. This starting point informs the way in which researchers analyze data, draw conclusions, and reflect back to predefined categories. Members of a certain community *do* often have a background, like professional, organizational, or in terms of experience, which researchers expect to inform their patterns of thought.

Although it is not necessary to account for the selection of participants (Watts and Stenner 2012), most researchers consciously select respondents. This is called a *purposive sample* (Kraak et al. 2014). As the debate under investigation will be most strongly demarcated by people who are very opinionated or knowledgeable, around 80% of the articles surveyed for this chapter begin by selecting respondents based on their professional, organizational, or personal background. Other studies make selections on the basis of the participants’ different countries, their role at different times in a certain policy process, etc. In essence, one only needs to select *enough* respondents who are *relevant* to the focus of the study, i.e. people who are able to assess the debate and hold distinct opinions (Van Exel and De Graaf 2005).

In terms of generalizing the results to a particular population based on the respondents’ characteristics, the methodology is limited. The conclusions only apply to those who took part in the study. Researchers do, however, consider the background or characteristics of the respondents by “eye-balling” a certain factor, and checking whether certain characteristics are more strongly present among the respondents who load on the factor.

Q methodology allows analysis of a small, non-representative sample of respondents called the P-sample (P here refers to *person*). Therefore, Q methodology can be considered a small-N method, both in terms of the number of participants as well as in the number of statements. Half of the articles in comparative policy analysis studied here engage around 20 to 40 participants, with 15 being the minimum (Nijnik, Nijnik, and Bizikova 2009) and 197 being the maximum (Morinière and Hamza 2012). Brown (1996) writes that many Q-methodology studies have around 30–50 respondents, which is considered adequate. Stephenson recommended a person–statement sample ratio of ≤ 2 (Stephenson 1953, cited in Mazur and Asah 2013: 82), which is a balance between a 2:1 and 3:1 ratio of statements to respondents (Cotton and Mahroos-Alsaiari 2015: 97; Webler, Danielson, and Tuler 2009).

3.4 Ranking by Participants

After purposefully selecting participants, the Q-sorting process can begin. Watts and Stenner (2012) call the Q-sorting process a “dynamic medium” in which subjectivity can be expressed, explained, and visualized. Stephenson considered sorting, and assessing how statements relate to each other an essential element of the Q methodology. The Q-sorting process is a ranking of statements – conducted either online or in person – in which statements are (often written on small cards) presented to the respondents in a random order (Van Exel and De Graaf 2005). These approaches are different, and have both advantages and disadvantages. Today, there are many options for conducting the Q-sorting procedure via a variety of online programs and tools. Using certain online tools (e.g. Q Method Software, FlashQ, and Q-assessor) means that the researcher cannot influence the thought process of the respondent, but also that he or she cannot ask why a respondent (dis)agrees with a particular statement. While face-to-face interviews are intensive and less time-efficient, they do provide the opportunity to study why certain discourses appear within certain groups of respondents (e.g. the mechanisms and explanations behind these discourses) and can be substantiated with actual quotes, examples, and explanations. Interviewees often come to understand their own opinions on the topic while they sort the items, and therefore live interviews can be very insightful. Watts and Stenner (2012) explain that the researcher should be cautious to interfere too much with the sorting process. Whenever a researcher explains a statement, he or she already attaches meaning to the card.

In interviews, an introductory page of a software tool, or at the beginning of an interview respondents can be instructed regarding how to rank the set of Q statements. Most researchers use a “structuring principle” or a condition of instruction, which is a concept created by Stephenson (see also Van Exel and De Graaf 2005). Here, Stephenson meant to emphasize the word “condition” in terms of preparing or conditioning the respondent for the sorting process. This can be a topic or a central question, e.g. your perception about joint commissioning in health and social care (Dickinson et al. 2014), or “consider what you think an ideal policy program for poverty would look like”. Since respondents rank statements in relation to other statements, this condition is important; it is the primary guideline helping the respondent order the statements. The instruction also includes an explanation about how the endpoints of the scale are labeled (e.g. “agree-disagree”, “in line with my ideas, not in line with my ideas”, etc.). It must be clear to the respondent under which condition he or she should rank the statements and according to which scale.

Most researchers instruct the interviewees to first sort the statements into three categories: agree, disagree, and neutral (see Van Exel and De Graaf 2005 for an excellent guideline on how to proceed with the interviews). This first step helps the respondent sort the many cards and proceed more easily with the following step. After sorting the cards in piles, the respondent begins to fill the grid. Ideally, the respondent should start from the extreme ends, as it is easiest to select one or two statements with which you most (dis)agree in comparison to others. Once all the statements are sorted the interviewer can discuss this sorting with the researcher to make certain the grid reflects her or his true opinions.

The researcher must be aware that Q methodology takes a “snapshot” of a certain topic at the specific time the research is conducted. Q methodology does not consider time, and researchers rarely execute time-series studies.

3.5 Analysis: Defining the Discourses

When the researcher has enough relevant respondents, the analysis can begin. PQmethod³ is the most frequently used analysis program. However, this is a DOS-based program, and more and more researchers are using SPSS or other statistical tools for principal component analysis. Another popular means is the Ken-Q package. The advantage of this program is that it is web-based, so no installation is needed.

The analysis reveals communalities among the respondents and looks for *saturation*. Saturation means placing as many respondents under a number of factors, so most respondents have subsequently a high degree of association with only one factor (McKeown and Thomas 2013: 52). Watts and Stenner (2012) explain this using the idea of a cake. The first piece taken from the whole cake (all the Q sorts) is the largest commonality. Afterwards, the remaining portion is analyzed and the largest degree of commonality is sought from within this “leftover” cake. For this reason, the first factor often displays the largest share of respondents and shows the highest explained variance. Most Q methodologists advise against referring only to statistical arguments alone to select the factors. However, there are some “statistical rules of thumb” that can be useful (from McKeown and Thomas 2013; Webler, Tuler, and Krueger 2001: 438; Webler et al. 2003: 112), which are explained below:

- Select only the factors that have **eigen values > 1** (Kaiser-Guttman criterion, Watts and Stenner 2012). This ensures that the researcher includes factors that explain a large proportion of the respondents.
- Select a factor if the **cross-product** of the two highest loadings is more than twice the standard error of the study (Humphrey’s rule, Brown 1980: 223, quoted in Watts and Stenner 2012).
- Select factors that have at least **two pure loadings**. A pure loading is a respondent that loads significantly on one factor, and only that factor. This will again ensure that the factors are relevant, and helps establish an explanation for why certain factors appear.
- Look for high **variance**, and choose individual factors that account for > 8% variance, and factors that account for a cumulative variance of > 30%. This latter point shows how much of the variance the factors explain together.
- Factors should **not be highly correlated**, as Q methodology seeks to understand different viewpoints in a debate. The inter-factor correlations should be < 0.33.

Other rules of thumb outlined in Watts and Stenner (2012) include the *scree test* (p. 108), which examines a drop in eigen values when plotted on a graph and *parallel analysis* (p. 109), which assesses whether the factors and eigen values truly present distinct factors or whether they appear to be random.

Most Q-methodological researchers explain that it is not sufficient to base factor selection on only statistical rules. As the method’s ultimate goal is to reveal subjectivity, one needs to go beyond a technically correct analysis (McKeown and Thomas 2013: 54). A few “non-statistical” rules (adapted from McKeown and Thomas 2013; Webler, Tuler, and Krueger 2001: 438; Webler et al. 2003: 112) include whether the factors are understandable and whether they are related to theory and practice, i.e. if they “make sense”.

After including all individual Q sorts, studying correlations and removing the largest communalities, the PQ method and Ken-Q package ask the researcher to indicate the Q sorts that load significantly on a particular factor. Watts and Stenner (2012: 198) refer to this as the

“eye-ball” analysis. To calculate the threshold significance level of an individual Q sort, one can use (Van Exel and De Graaf 2005: 19):

$$1.96 \times (1/\sqrt{\# \text{ of statements}}) = \text{significance level of } p < 0.05$$

$$2.58 \times (1/\sqrt{\# \text{ of statements}}) = \text{significance level of } p < 0.01$$

$$3.29 \times (1/\sqrt{\# \text{ of statements}}) = \text{significance level of } p < 0.001.$$

If, based on the significance levels above, the researcher has established which Q sorts load on a particular factor he or she can “flag” these and commence with retrieving the factors. Afterwards, the analysis package produces an overview. This overview provides information about *pure loadings*, i.e. the respondents who load on a particular factor. The respondents that load most highly on a particular factor are called “exemplars”, as they are most clearly correlated with the factor (Willis and Jeffares 2012: 547). This informs the researcher about which specific respondents may be able to offer an illustration or a quotation about a particular factor. It also helps the researcher understand *why* this particular respondent is associated with a particular factor, as he or she can refer to the study’s participant selection criteria and interview material to understand the discourse.

The analysis also produces *characterizing statements* (Van Exel and De Graaf 2005), which are statements that are scored highly or low in a certain factor. These statements help reconstruct the discourse, and together with the distinguishing statements are key to understand the discourse. *Distinguishing statements* are also a helpful element provided in the overview created by the analysis package. These statements are significantly different compared to other factors. The distinguishing statements show the basic statements on which the discourses differ and as such function as a type of “signpost” highlighting significant differences. Although they are not always on the extreme ends of the scale like the characterizing statements, distinguishing statements are important for understanding a certain perspective.

Finally, the overview provides *consensus statements*. These statements are not placed significantly differently across factors. However, we do not know *why* all respondents place the statement in a particular column in the grid. In general, all that is known about consensus statements is that most respondents placed the statement in more or less the same sorting position.

Most researchers begin to get an understanding of the results by studying the extreme statements and distinguishing statements, as these represent the statements on which the respondents feel most strongly. Factor interpretation can be endless because there are different readings and meaning attached to the factors (Watts and Stenner 2005), and quotes and interviews can help clarify the factors more precisely.

3.6 Drawing Conclusions Based on the Discourses

After defining and describing the statements that belong to a certain factor, the researcher attempts to label the discourse and understand it. Most researchers try to interpret the factors, i.e. try to understand: “what do these respondents think about this debate?” and “how can we understand their viewpoints?”

In interpreting and defining these factors there are various ways to proceed. Some researchers delve back into their respondent selection and begin analyzing whether the factor configuration relates to the predefined groups. As is clear from most of the articles considered in this chapter, the discourses often become apparent irrespective of the predefined groups. So,

what can be concluded after a Q-methodological study? Explaining why certain discourses are present is not the end goal of the method. Q methodology is in essence a descriptive method, as it reveals communalities. Some Q methodology researchers try to outline areas of consensus and conflict when comparing factors. Other Q methodology researchers analyze a single factor more thoroughly, and see a factor as an “advocacy coalition” (Wolsink 2004: 2675), i.e. as a group that holds a shared belief regarding certain aspects of the discourse. If one wants to explain why a discourse arises, the analysis can be enriched through focus groups or by returning to the interviews and exploring them in more depth.

Q methodology strives for conceptual generalization (Watts and Stenner 2012). As the clusters represent particular thought patterns on a particular topic, Q methodology can enrich our knowledge about particular ideas and topics and outline the multifaceted nature of certain concepts. In terms of what the methodology delivers, Q methodology deepens our understanding of policy realities, as it not only illustrates the majority perspective, but also minority ones. Q methodology looks for a rich explanation for policy design struggles, and “by losing generalization for complexity, this qualitative approach eventually leads to another way of theory building and testing” (Peters, Fontaine, and Mendez 2018: 137).

Many New Public Governance (NPG) and public value scholars urge researchers to think about values and perceptions as existing in parallel with facts and figures. Furthermore, some researchers explain that values must be considered when designing a new policy (tool) around a specific theme (Pelletier et al. 2000). Q methodology can help to both guide this search and design better policies because it makes cognitive frameworks more explicit. New policy proposals that acknowledge both minority and majority viewpoints may help to overcome a tardy process later. In Q methodology, the factors often reveal that what works for one group of respondents (established by a certain factor), will be counterproductive for others. By illuminating these significant clusters, Q methodology can provide important empirical information for the creation of (new) policy programs and customization of old ones. For instance, Vugteveen et al. (2010) demonstrate how a Q analysis can be very beneficial for developing planning scenarios, while Dickinson et al. (2014) illustrate how governments can approach joint commissioning using Q methodology.

4. CONCLUSION

Policy design and modification is a multi-dimensional matter that raises many questions which policy analysts are interested in answering (Peters, Fontaine, and Mendez 2018). Such questions include: “What are the underlying viewpoints that informed a policy program?” “What range of views are there, and where do they overlap and conflict?” In this chapter, Q methodology is presented as one of the methods that can be included in the policy analysis toolbox to conduct a comparative analysis. This small-N method allows studying a small, non-representative sample of respondents and a small sample of statements (a subset) derived from the debate around a particular issue. This method reveals significant differences in people’s viewpoints. The method is not particularly suitable for understanding causal mechanisms and does not incorporate time issues well. The method’s strengths are providing insight into both dominant and minority viewpoints, as well as points of conflict and overlap in particular policy design issues. These assets can provide an empirical basis when creating new policy programs or reconfiguring existing ones.⁴

NOTES

1. Current release is PQMethod 2.35 with PQROT 2.0 (10-Nov-2014): <http://schmolck.org/qmethod/#PQMethod>, accessed January 29, 2020.
2. <https://shawnbanasick.github.io/ken-q-analysis/>, accessed January 29, 2020.
3. Current release is PQMethod 2.35 with PQROT 2.0 (10-Nov-2014).
4. This chapter greatly benefited from the feedback of Dr. Amanda Wolf, of the Victoria University of Wellington, and from the feedback received during the International Public Policy Association (IPPA) International Workshops on Public Policy at the University of Pittsburgh, June 26–28, 2018.

REFERENCES

- Allison, G. T. (1971). *Essence of Decision: Explaining the Cuban Missile Crisis*. Boston, MA: Little, Brown & Company.
- Ansell, C. and Gash, A. (2007). Collaborative governance in theory and practice. *Journal of Public Administration Research & Theory* 18 (4): 543–71.
- Brown, S. R. (1974). The composition of microcosms. *Policy Sciences* 5 (1): 15–27.
- Brown, S. R. (1980). *Political Subjectivity: Applications of Q Methodology in Political Science*. New Haven, CT: Yale University Press.
- Brown, S. R. (1993). A primer on Q methodology: operant subjectivity. *Operant Subjectivity: The International Journal of Q Methodology* 16 (3/4): 91–138.
- Brown, S. R. (1996). Q methodology and qualitative research. *Qualitative Health Research* 6 (4): 561–7.
- Brown, S. R. (2008). Q methodology. In L. M. Given (Ed.), *The Sage Encyclopedia of Qualitative Research Methods*. Thousand Oaks, CA: Sage, 700–704.
- Cotton, M. D. and Mahroos-Alsaiani, A. (2015). Key actor perspectives on stakeholder engagement in Omani Environmental Impact Assessment: an application of Q-methodology. *Journal of Environmental Planning and Management* 58 (1): 91–112.
- Cuppen, E. (2012). Diversity and constructive conflict in stakeholder dialogue: considerations for design and methods. *Policy Sciences* 45 (1): 23–46.
- Curry, R., Barry, J., and McClenaghan, A. (2013). Northern visions? Applying Q methodology to understand stakeholder views on the environmental and resource dimensions of sustainability. *Journal of Environmental Planning and Management* 56 (5): 624–49.
- Davies, B. B., Blackstock, K., and Rauschmayer, F. (2005). “Recruitment”, “composition”, and “mandate” issues in deliberative processes: should we focus on arguments rather than individuals? *Environment and Planning C: Government and Policy* 23 (4): 599–615.
- de Wijs, L., Witte, P., and Geertman, S. (2016). How smart is smart? Theoretical and empirical considerations on implementing smart city objectives – a case study of Dutch railway station areas. *Innovation: The European Journal of Social Science Research* 29 (4): 424–41.
- Dickinson, H., Jeffares, S., Nicholds, A., and Glasby, J. (2014). Beyond the Berlin Wall? Investigating joint commissioning and its various meanings using a Q methodology approach. *Public Management Review* 16 (6): 830–851.
- Dimitrova, A. and Kortenska, E. (2017). What do citizens want? And why does it matter? Discourses among citizens as opportunities and constraints for EU enlargement. *Journal of European Public Policy* 24 (2): 259–77.
- Dryzek, J. S. and Berejikian, J. (1993). Reconstructive democratic theory. *American Political Science Review* 87 (1): 48–60.
- Eeten, M. J. G. van (2001). The wider implications of the Netherlands civil aviation controversy. *Policy Analysis* 20 (3): 391–414.
- Howlett, M. (2009). Governance modes, policy regimes and operational plans: a multi-level nested model of policy instrument choice and policy design. *Policy Sciences* 42 (1): 73–89.
- Jeffares, S. and Skelcher, C. (2011). Democratic subjectivities in network governance: a Q methodology study. *Public Administration* 89 (4): 1253–73.

- Kraak, V. I., Swinburn, B., Lawrence, M., and Harrison, P. (2014). A Q methodology study of stakeholders' views about accountability for promoting healthy food environments in England through the Responsibility Deal Food Network. *Food Policy* 49: 207–18.
- Mazur, K. E. and Asah, S. T. (2013). Clarifying standpoints in the gray wolf recovery conflict: procuring management and policy forethought. *Biological Conservation* 167: 79–89.
- McKeown, B. and Thomas, D. (2013). *Q Methodology*. Series: Quantitative Applications in the Social Sciences. Thousand Oaks: SAGE Publications.
- Morinière, L. C. E. and Hamza, M. (2012). Environment and mobility: a view from four discourses. *Royal Swedish Academy of Sciences* 41 (8): 795–807.
- Nekola, M. (2012). Pragmatists, prohibitionists and preventionists in Czech drug policy. *Central European Journal of Public Policy* 6 (2): 56–83.
- Nijnik, M., Nijnik, A., and Bizikova, L. (2009). Analysing the development of small-scale forestry in Central and Eastern Europe. *Small-Scale Forestry* 8 (2): 159–74.
- Nikraftar, T. (2013). The conceptions about managing coordination in Iranian governmental machinery: A Q methodology study. *Quality and Quantity* 47 (5): 2577–96.
- Ockwell, D. G. (2008). “Opening up” policy to reflexive appraisal: a role for Q methodology? A case study of fire management in Cape York, Australia. *Policy Sciences* 41 (4): 263–92.
- Pelletier, D. L., Kraak, V., McCullum, C., and Uusitalo, U. (2000). Values, public policy, and community food security. *Agriculture and Human Values* 17 (1): 75–93.
- Peritore, N. P. and Peritore, A. G. (1990). Brazilian attitudes toward agrarian reform: a Q-methodology opinion study of a conflictual issue. *Journal of Developing Areas* 24 (3): 377–406.
- Peters, B. G. (1998). Managing horizontal government: the politics of co-ordination. *Public Administration* 76: 295–311.
- Peters, B. G., Fontaine, G., and Mendez, J.-L. (2018). Introduction: substance and methods in the comparative study of policy change. *Journal of Comparative Policy Analysis: Research and Practice* 20 (2): 133–41.
- Peters, B. G. and Hoornbeek, J. (2005). The problem of policy problems. *Journal of Comparative Policy Analysis: Research and Practice* 7 (4): 349–70.
- Ramlo, S. E. and Newman, I. (2011). Q methodology and its position in the mixed-methods continuum. *Operant Subjectivity: The International Journal of Q Methodology* 34 (3): 172–91.
- Ramlo, S. E. and Newman, I. (2015). Theoretical significance in Q methodology: a qualitative approach to a mixed method. *Research in the Schools* 22 (1): 172–91.
- Rittel, H. W. J. and Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences* 4 (2): 155–69.
- Scharpf, F. W. (2000). Institutions in comparative policy research. *Comparative Political Studies* 33 (6–7): 762–90.
- Stenner, P., Watts, S., and Worrell, M. (2007). Q methodology. In C. Willig and W. Stainton Rogers (Eds.), *The SAGE Handbook of Qualitative Research in Psychology*. London: Sage, 215–39.
- Termeer, C., Dewulf, A., and Breeman, G. (2013). Governance of wicked climate adaptation problems. In J. Knieling and W. L. Filho (Eds.), *Climate Change Governance*. Berlin: Springer, 27–41.
- Van Exel, J. and De Graaf, G. (2005). Q methodology: a sneak preview. *Social Sciences* 2: 1–30.
- Vugteveen, P., Lenders, H. J., Devilee, J. L., Leuten, R. S., Van Der Veeren, R. J., Wiering, M. A., and Hendriks, A. J. (2010). Stakeholder value orientations in water management. *Society and Natural Resources* 23: 805–821.
- Watts, S. and Stenner, P. (2005). Doing Q methodology: theory, method and interpretation. *Qualitative Research in Psychology* 2: 67–91.
- Watts, S. and Stenner, P. (2012). *Doing Q Methodological Research: Theory, Method & Interpretation*. London: Sage.
- Webler, T., Danielson, S., and Tuler, S. (2009). *Using Q Method to Reveal Social Perspectives in Environmental Research*. Greenfield, MA: Social and Environmental Research Institute.
- Webler, T., Tuler, S., and Krueger, R. (2001). What is a good public participation process? Five perspectives from the public. *Environmental Management* 27 (3): 435–50.
- Webler, T., Tuler, S., Shockey, I., Stern, P., and Beattie, R. (2003). Participation by local governmental officials in watershed management planning. *Society and Natural Resources* 16 (2): 105–21.

- Willis, M. and Jeffares, S. (2012). Four viewpoints of whole area public partnerships. *Local Government Studies* 38 (5): 539–56.
- Wlezien, C. (2004). Patterns of representation: dynamics of public preferences and policy. *Journal of Politics* 66 (1): 1–24.
- Wolsink, M. (2004). Policy beliefs in spatial decisions: contrasting core beliefs concerning space-making for waste infrastructure. *Urban Studies* 41 (13): 2669–90.

APPENDIX

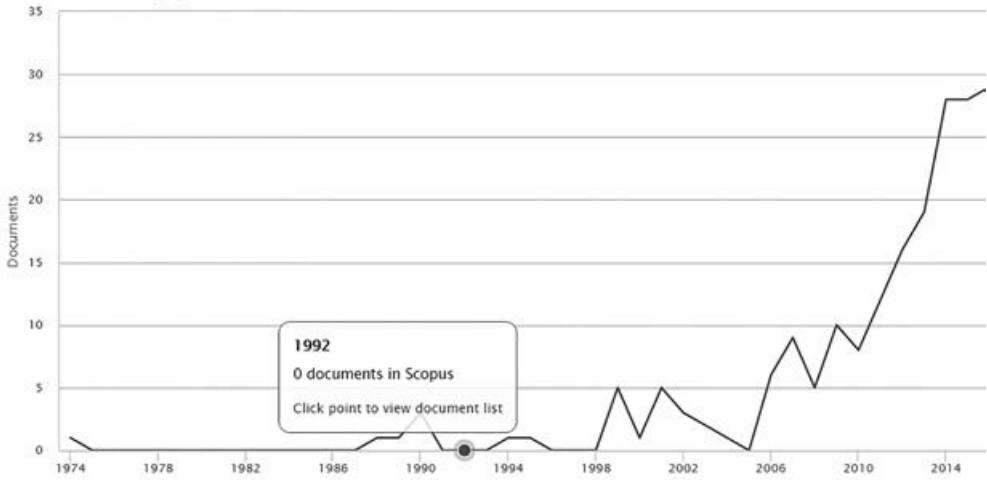
Although Q methodology is not a recently developed method, it seems that the value it brings has gained more recognition in recent years. In public administration and more specifically in studying public policy, Q methodology is a useful method to apply. Considering the ideas put forward in the introduction, this method appears to support the comprehension of wicked policy problems and stakeholder frames.

A literature study was carried out in order to obtain insight into the Q-methodological applications in (comparative) policy analysis. Based upon a systematic search via the search engine Scopus, 40 articles were studied to determine how they apply Q methodology. Two keywords were entered into the search engine: Q-metho* (because some authors refer to Q-method or Q methodology) AND policy. This led to a long list of 195 articles and five book chapters. In refining this list, all book chapters were excluded from the outset. The largest group of articles in the result list was from the social sciences (n = 119), followed by environmental science (n = 91). Some articles were part of multiple categories. The main journals represented in the results were *Ecological Economics* (n = 10), *Policy Sciences* (n = 8), and the *Journal of Policy Analysis and Management* (n = 5). The oldest article is from 1974 and the second oldest is from 1988 (see Figure 19A.1). In the period from 1999 to 2005 the average number of articles applying Q methodology in public policy per year was 3, while by 2016 this number had grown tenfold (see Figure 19A.1). The abstracts from the 195 articles were scanned for eligibility for the next round of analysis based on the following criteria:

- Only articles that have a *comparative element* were selected. Q methodology is in itself comparative, however, only publications with a comparative focus, i.e. a specifically mentioned comparative design, were selected. This comparative element could relate to respondents, policy issues, respondent background, or case (being a country, county, organization, etc.).
- Articles that address *public matters* were selected; those about public goods and services, and policy decision making and implementation. Moreover, public organizations or civil servants had to be part of the study to some extent, in order to be part of the selection.
- Articles addressing policy (making) *as an object of study* were selected. Articles mentioning general ideas that could possibly be used for policy improvement were omitted.
- Conceptual and theoretical papers were not considered. Only empirical articles were selected.

After the initial round of reviewing abstracts and determining whether the articles were eligible for analysis, 108 studies remained. These 108 were reviewed more closely and ultimately 40 articles were selected for this chapter.

Documents by year



Source: Output from Scopus, retrieved June 2017.

Figure 19A.1 Applications of *Q* methodology in policy studies

20. Using the Narrative Policy Framework in comparative policy analysis

Aaron Smith-Walter and Michael D. Jones

1. INTRODUCTION

In an influential article published in the *American Political Science Review* in December 1970, Giovanni Sartori observed that a methodological problem lay at the heart of the approach undertaken by comparative political studies as the field expanded outside its traditional geographic boundaries. Sartori noted that political scientists had inherited numerous concepts from previous generations which led to the emergence of a type of scholarship which tended to uncritically apply categories developed in one context to those of another (Sartori 1970: 1033–4). Sartori was critical of this tendency to engage in “conceptual stretching”, whereby existing descriptions of phenomena that were articulated in particular political systems and traditions (primarily Western, see Wong 2014) are applied to systems where their fit is not exact. This, he contends, results in comparative politics generating “vague, amorphous conceptualizations” (Sartori 1970: 1034), which result in the production of findings that are so obvious as to be of little import.

For Sartori, the appropriate solution to conceptual stretching was for comparative scholars to develop concepts which were not only universally applicable and empirically measurable, but also able to move up and down the “ladder of abstraction” in order to link research conducted at various levels of analysis. The “ladder of abstraction” was a metaphor adopted by Sartori to emphasize that universal categories must be amenable to systematic sub-categorization to reflect meaningful differences in local manifestations of the universal category. Concepts that can “travel”, he argued, are vital as they allow findings produced in one study to meaningfully inform work undertaken at other levels and in different contexts.

In this chapter we argue that the Narrative Policy Framework (NPF) provides comparative public policy scholars valid categories that can address Sartori’s dilemma, as the NPF’s concepts are optimal for “traveling” across different contexts. We lay the foundation for this argument by first describing the components of the NPF. We then explore two reasons the NPF can be fruitfully applied to comparative policy studies: the universal applicability of narrative to human cognitive and communicative endeavors and the ability of the NPF to integrate findings from qualitative case studies with quantitative statistical testing inside the scaffolding of a single framework. The chapter closes with discussions of potential limitations of the NPF for comparative studies.

2. THE NARRATIVE POLICY FRAMEWORK

Named in 2010 (Jones and McBeth 2010), the NPF was formulated by Mark K. McBeth, Elizabeth A. Shanahan, and Michael D. Jones working from the idea that since stories are

important for individual human communication and cognition they were likely important to groups of humans engaged in the “processes, outcomes, implementation, and designs of public policy” (Jones, McBeth, and Shanahan 2014: 1). The NPF takes as its starting point that narrative is, as Walter R. Fisher argues, “meaningful for persons in particular and in general, across communities as well as cultures, [and] across time and place” (1984: 8). Building on this understanding, the NPF approaches the study of public policy with an objective epistemology embracing the tenets of science, coupled with a subjective ontology focusing on the centrality of narrative meaning making in politics (Jones and Radaelli 2015). This combination allows the NPF to operate in the scientific realms of both qualitative and quantitative methods (e.g. Shanahan et al. 2017) and also embrace interpretivist approaches to public policy (Gray and Jones 2016). This orientation places it squarely in the tradition of the policy sciences, which values acknowledging multi-disciplinary approaches with explicitly value-oriented research perspectives (DeLeon 2006: 40–41).

The NPF conceives of policy narratives as possessing two distinct parts, *form* and *content*. Narrative form is the way a story structures four different elements that are foundational to narrative communication, which are *setting*, *plot*, *characters*, and *a moral* (Shanahan et al. 2017). These elements can be identified, quantified, generalized, and, of course, interpreted. Such operations facilitate comparison across contexts.

For instance, every policy narrative possesses a setting, best understood as the stage on which the characters act out their roles. The setting consists of such things as existing laws, facts held as common knowledge, the economic situation, issue frames (Shanahan et al. 2018) and other policy consequential information (Jones, McBeth, and Shanahan 2014).

The plot is the way that a policy narrative arranges interactions across time and space between the characters and the setting. Perhaps best thought of as a narrative arc, plot has been operationalized in different ways within the NPF, including blame attribution (Crow and Berggren 2014), but most studies have relied on some variant of Stone’s (2012) conception of policy plots (e.g. Jorgensen, Song, and Jones 2018). As with all of its narrative concepts, the NPF encourages innovation and the borrowing from other fields and modes of inquiry and is amenable to other potential approaches to plot (for instance, see Booker 2004).

The NPF initially recognized three types of characters: the *villain* who creates harm, the *victim* who is (or will be) harmed, and the *hero* who promises to alleviate the damage or threat (Shanahan et al. 2017). It is important to note that in the eyes of the narrator, characters need not be human, as the environment, cherished values, and existing laws can be characters as well (although see Weible et al. 2016). The addition of new characters is not precluded by the framework as shown by the recent introduction of new characters, such as “Charismatic Experts” (see Lawton and Rudd 2014) “Allies”, “Opponents”, “Perpetrators” (Merry 2016), and “Beneficiaries” (Huda 2018b; Weible et al. 2016) which were added to account for particularities in the policy narratives in varied contexts. It is also possible for narratives to include multiple characters of the same type.

The final element of narrative form is the *moral*, which is typically understood as a policy solution or a call to action of some sort. This element is linked to the action that must be taken by the hero to thwart the villain and/or aid the victim. These four narrative elements all play important roles; however, the presence of all four in a single narrative is not a necessary condition for a policy narrative to exist. Instead, the NPF holds that a “policy stance or judgement of policy-related behavior + story character = policy narrative” (Shanahan et al. 2013: 457).

Narrative content deals with the minutiae of the specific policy issue, its storied characters, the nuances of the setting, the particulars of a certain characterization of the problem at hand, etc., and is thus context-dependent and some would argue, impossible to generalize (especially when one considers the variability associated with the interaction of narrative communication and individuals) (see, for example, Dodge 2015's critique of the NPF). In other words, a story populated with climate change content is simply different from a story populated with gun control content, and how those stories are interpreted by individuals varies quite a bit. In the NPF we call the *sui generis* nature of narratives the problem of narrative relativity (see Jones, McBeth, and Shanahan 2014). To moderate the problem of narrative relativity the NPF argues that one can employ generalizable belief systems or strategies. For example, robust academic theories attempting to capture *belief systems* such as cultural worldviews (e.g. Jones 2014), ideology (Clemons, McBeth, and Kusko 2012), or orientations toward citizenship (McBeth, Lybecker, and Garner 2010) can be identified within policy narratives and can serve as meaning anchors for elements such as characters or specific pieces of the setting. They are meaning anchors in the sense that some non-trivial component of the population interacting with them will have systemic and general reactions to said anchors. For example, Hillary Clinton will illicit general – but quite different – reactions from both conservatives and liberals in the US. Similarly, a study by the IPCC is likely to have generalizable ideological qualities when used in a climate change narrative within the US. But general belief systems are often insufficient, and one may find a need for capturing more localized and nuanced beliefs germane to the policy problem at hand. For example, McBeth, Shanahan, and Jones operationalized federal (New West) and state/local level (Old West) belief systems in the US associated with Yellowstone National Park by observing the difference between heroic portrayals of governmental actors at different levels of government (2005).

Within the NPF, narrative strategies are operationalized similarly to belief systems. The researcher simply looks for systematized patterns of behavior by actors deploying narratives. One of the most utilized strategies explored by the NPF is the devil-shift (Sabatier, Hunter, and McLaughlin 1987), which considers the frequency with which villains appear in a policy narrative. This allows the NPF to explore how certain narratives may try to persuade or mobilize their audience by creating a palpable threat to something or someone they hold dear (Smith-Walter et al. 2016). Other strategies studied in NPF research include attempts to manage the scope of conflict (Schattschneider 1960), blame attribution (Stone 2012), and heresthetics (Riker 1986).

In short, the NPF's structural approach to the study of narrative allows a researcher to identify universal narrative from categories such as settings, plots, characters, and policy solutions as well as generalizable content through belief systems and strategies. We hold that these generalizable narrative elements coupled with an objective epistemology and subjective ontology have the potential to contribute valuable tools to the study of comparative public policy.

Generalizability is, however, only half of what Sartori's call requires. He also stresses the need for concepts to be applicable to multiple levels of abstraction to produce "sets of propositions which either reinforce or contradict the propositions of the neighboring levels" (Sartori 1970: 1053). Here, the NPF is positioned to meet the challenge, as it posits micro-, meso-, and macro-levels of analysis which operate simultaneously in the policy process and are associated with their own theoretically derived hypotheses (see Shanahan et al. 2017: 179–98). The micro-level concerns itself with how individuals process and deploy narrative commu-

nications (Shanahan et al. 2017). The micro-level has predominantly relied on experimental methods and surveys (Jones and Song 2014; Lybecker, McBeth, and Stoutenborough 2016).

The meso-level of the framework focuses on policy narratives composed and disseminated by groups and coalitions within a given policy subsystem. This level emphasizes how policy narratives work to build, manipulate, destroy, and rebuild coalitions of actors in and across policy subsystems (Shanahan, Jones, and McBeth 2011). It also explores how policy beliefs, held by groups, vary over time (McBeth et al. 2010; Shanahan et al. 2013).

The macro-level is relatively underdeveloped, but it is conceived as an examination of “cultural and institutional policy narratives that condition and permeate social bodies over long periods of time to determine how such narratives shape public policy” (Jones, McBeth, and Shanahan 2014: 19). While few studies have sought to study narratives at the macro-level, Ney (2014) used Cultural Theory (Douglas and Wildavsky 1983; Thompson, Ellis, and Wildavsky 1990) to identify three cultural “grand narratives” to analyze competing definitions of “social innovation” and its relationship to change in public and non-profit organizations. To the extent that macro-level narratives can be identified across numerous policy areas and nations, the macro-level of the NPF could prove useful in comparatively exploring the role of cultural values and biases in the crafting of public policy.

While quantitative measurement of narrative components is often emphasized because of the NPF’s objective epistemology, the framework’s ontology is decidedly social constructivist (Jones, McBeth, and Shanahan 2014), allowing its extension into qualitative approaches, as evidenced by Gray and Jones’s study published in 2016. In their study on campaign finance narratives in the US, Gray and Jones (2016) found two predominant values, equality and freedom of expression. Specific victims, villains, and heroes differed across groups as did the moral of the story. This study argued that by exchanging “commonly recognized qualitative standards” (Gray and Jones 2016: 198) for traditional quantitative standards (i.e. scientific), the NPF was able to accommodate interpretive and qualitative modes of inquiry, which are often central to effective case studies.

We will next argue that narratives are: (1) universal and (2) structurally amenable to empirical study. This is a necessary step, since our view that the NPF meets Sartori’s call carries with it the implication that the categories are applicable at multiple levels of analysis.

3. UNIVERSALITY OF NARRATIVE

One of the primary difficulties of comparative policy research is dealing with varied contexts. In simple terms, countries, cultures, and people, in disparate locations, vary substantially on any number of dimensions and identifying theoretical concepts that “travel” well between these contexts is a perpetual problem. As Sartori (1970) notes, most work on methodology in the social sciences “has little if anything to share with the crucial concern of ‘methodology’, which is a concern with the logical structure and procedure of scientific enquiry. In a very crucial sense there is no methodology without *logos*, without thinking about thinking” (Sartori 1970: 1033). This has tended to result in an unhappy situation where concepts used in comparative policy research become susceptible to “conceptual stretching”, and a general loss of connection to empirical observation (Sartori 1970: 1035).

The NPF’s approach to “thinking about thinking” entails a step back from focusing on the impact of institutions on public policy outcomes (Gupta 2012) and instead turns the spotlight

on the manner in which human understanding, coordination, and competition are continuously (re)structured by narratives. To do so the NPF understands narrative as foundational to human cognition and communication, a notion that is not only supported by extant NPF research, but also supported in numerous academic literatures. For example, political scientists have observed narrative to be an organizing cognitive heuristic for people dealing with new information (e.g. Berinsky and Kinder 2006; Jones and Song 2014). Marketing researchers have found that narrative sales techniques are more effective among consumers with limited experience with a product or service than traditional price point techniques (Mattilla 2000). Communication scholars have found that narratives that are more immersive are also more persuasive and memorable (Green and Brock 2000). Rhetorical scholars have argued that narrative is a foundational form of human understanding and communication (Fisher 1984, 1985, 1989; McClure 2009). And literary scholars have argued that narrative is fundamental to the way in which human beings organize and make sense of their experiences (Herman 2009).

Some of the most compelling findings for the universal importance of narrative emerge from neuroscience, where researchers have located narrative centers in the brain (Mar 2004) and found that damage to these areas is more problematic than other areas of the brain such as kinesthetic or linguistic centers. Damage to narrative centers, such as those found in long-term alcoholics or those with Alzheimer's disease, actually result in individuals being unable to understand their own identities as they cannot position themselves in time and space relative to other actors in their lives (Ash et al. 2007). Additional research on human neurological responses to narratives found that a video containing an emotionally compelling narrative resulted in the production and release of more oxytocin by tested individuals than a control video without the emotional content (Zak 2015). Oxytocin is a neurochemical which is produced by humans when they experience a feeling of trust and is correlated with increased reciprocity (Zak, Kurzban, and Matzner 2005). Lin et al. (2013) found that male subjects who received injections of synthetic oxytocin prior to the viewing of public service announcements (PSAs) donated over 50% more money to alleviate the public problem than the control group who received saline. In short, narrative is a universal characteristic of human biology, and most certainly not a product limited to any one particular culture. As Sugiya notes:

If narrative were a cultural invention, one would expect to find evidence of its having spread by contact and of its being extremely elaborated in some cultures and absent in others. This is not the case. Although subject matter is often borrowed from other cultures, the practice of storytelling itself emerges independently among even the most isolated peoples. (2001: 222)

Having presented our case for narrative's universality, we now turn our attention to three vital aspects of the universality of narrative: narrative rationality, narrative persuasion, and narrative and behavior.

3.1 Narrative Rationality

Some have taken the identified importance of narrative to indicate that narrative presents its own form of universal rationality, independent from other concepts such as instrumental or communicative rationality (Bennett and Edelman 1985; Fisher 1989: 57; Jones, McBeth, and Shanahan 2014). The NPF adopts this *Homo narrans* model of rationality (Shanahan et al. 2017), built upon ten postulates related to cognition and decision making (see Table 20.1).¹

Table 20.1 *Homo narrans* model of the individual

Postulates	Impact on cognition and decision making	Attribution
Bounded rationality	Constraints on information gathering lead to selection of sub-optimal solutions – e.g. satisficing.	Simon 1947
Heuristics	Use of shortcuts such as past experience, disciplinary standards, specialized training, etc. to make decisions in place of a rigorous calculation of all possible payoffs associated with all possible choices.	Kahneman 2011
Primacy of affect	Positive and/or negative emotive responses precede and influence the direction of rational thought.	Lodge and Taber 2005, 2007
System 1 & System 2 cognition	System 1 cognition relies on innate biological responses and habituation, to make multiple decisions at one time with limited energy expended, most “thinking” is of this type. System 2 cognition deals with complex problem-solving and is time and resource-intensive, undertaken less frequently and is performed in a serial fashion.	Kahneman 2011
Hot cognition	New concepts are affectively organized in a manner consistent with existing mental model of the world.	Morris et al. 2003
(Dis)Confirmation bias	Information that comports with existing belief is more easily and quickly assimilated than that which contradicts established understanding.	Taber and Lodge 2006
Selective exposure	Individuals reduce discomfort associated with challenges to the accuracy of their mental model by selecting information outlets that confirm their existing point of view.	Stroud 2008
Identity-protective cognition	An individual’s self-concept is valuable to them and they work to insulate it from challenge.	Kahan et al. 2007
Primacy of groups and networks	The social ties in which people are embedded function to condition affect toward political ideas and programs.	Kahan and Braman 2006
Polkinghorne	Narrative functions as a filter between the cacophony of inputs and the human need to organize events into a coherent series of related events.	Polkinghorne 1988

Source: Adapted from Shanahan et al. (2017: 181–3).

Homo narrans is an important tool for comparative scholars interested in exploring how institutions are created, sustained, and even subverted by the actors’ subject to their putative constraints. As Schmidt (2010) notes, much work from the traditions of rational choice institutionalism and historical institutionalism has suffered since they have assumed that positions held by individuals, in institutions, are synonymous with power. However, the actual power exercised by individuals is variable since, “ideas and discourse about how they can and should wield power ... may reinforce or undermine the power they derive from their position, depending upon the responses of their audience to their stated purposes” (Schmidt 2010: 18). A good example of this in contemporary US politics can be seen in the debates surrounding the Affordable Care Act (ACA) where the necessity of reforming the system of healthcare in the US was agreed upon by both political parties, but President Obama’s plan for extending health insurance to all citizens was vehemently opposed by many conservatives who saw it as federal overreach into the liberty of individuals. Obama used narratives to counter these concerns, highlighting the current state of American healthcare which left millions uninsured. Obama did not ask his audience to perform a cost–benefit analysis to determine whether or not they would save money, but instead appealed to characters that audiences could identify with and values that they could support (Smith-Walter 2018). The rationality of narrative can thus be understood to be one that takes seriously the power of symbols to advance political goals (Edelman 1985) through persuasion (and attention, see Peterson 2018), and is most certainly portable across contexts.

3.2 Narrative Persuasion

Lodge notes that “the logic of comparative public policy is driven by the search for determinants of public policy” (2007: 275). It is our contention that along with traditional variables associated with comparative public policy, such as ideology, institutional structures, voting behavior, public bureaucracy, and underlying social, economic, and cultural factors, policy narratives should be considered foundational to the study of public policy in democratic societies (Roe 1994; Schmidt 2010). This is because policy making is, at its base, concerned with convincing the right audience to support action intended to address a public problem (Majone 1989; Radaelli, Dunlop, and Fritsch 2013). Such support is by definition universal, and the persuasion therein likely influenced by its narrative nature. Below, and in the interest of guiding scholars to research illuminating narrative persuasion, we point out a few salient studies.

As the study of politics has moved from “government” to “governance” (Goodin, Rein, and Moran 2006: 12) the need to understand how persuasion functions to promote the successful passage and implementation of particular policy proposals across disparate contexts is a vital undertaking for scholars in the comparative tradition. For instance, using a quasi-experimental approach, Ertas (2015) found that the presentation of a narrative skeptical of charter schools as a solution to failings in conventional public schools caused a greater influence on support for charter schools than did a narrative that was approving of charter schools. Additionally, this change was more pronounced in individuals with a low level of familiarity with the charter school debate (Ertas 2015: 443). Jones and Song (2014) found that individuals presented with climate change narratives that were congruent with their pre-existing cultural worldviews (Thompson, Ellis, and Wildavsky 1990), were more likely to remember information in ways that reflected the structure of the narrative presented to them. One last illustration on the power of narrative components to persuade is displayed in the work of Jones (2014). Using an experimental survey design assessing US orientations toward climate change, Jones found that positive affect for the hero led to increased support for the arguments and assumptions embedded in the narrative. This experiment was replicated with Norwegian citizens, with similar results (Jones, Fløttum, and Gjerstad 2017). Taken together, these insights indicate that the way policy narratives are formulated is likely to have an impact on an audience. Exactly what impact is, is the topic to which we now turn.

3.3 Narrative and Behavior (Strategy and Mobilization)

If narratives are persuasive, it follows then that they also likely shape behavior. Recent research performed by Smith et al. (2017) with the Agta, a hunter-gatherer society in the Philippines, shows that narratives are key to developing social action. The researchers asked Agta adults to tell them stories that would normally be relayed to children. The stories transmitted values by using characters such as the sun and the moon, or ants and birds, to teach lessons “relevant to coordinating behavior in a foraging ecology, such as cooperation, sex equality and egalitarianism” (Smith et al. 2017: 1). The study also found that camps with a higher proportion of skilled storytellers exhibited higher levels of cooperation, and skilled storytellers had greater reproductive success and captured a larger portion of resources than less skilled storytellers.

The mobilization of support, whether from the public at large, or specific stakeholder groups, is a key consideration in many policy proposals. For instance, Blair and McCormack

found that a liberal newspaper in Colorado more frequently employed environmentalists as heroes in policy narratives on hydraulic fracturing, while the conservative paper featured oil and gas companies as their heroes (2016). This association between political views and groups or institutions as heroes may then result in positive affect, which increases support for policy solutions championed by said heroes (e.g. Jones 2014). As such, narratives may effectively encourage individuals to engage in behaviors such as donating money, communicating with public officials, casting votes for particular candidates, or even taking to the streets in acts of protest or civil disobedience.

Narratives can also serve to limit engagement in politics by downplaying the importance of an issue or arguing that the problem is under control. This strategy of trying to limit the scope of conflict is another important way that policy narratives can shape behavior. Additionally, narratives can frame problems in ways that seem to lead to the solution that is preferred by the authoring organization (Radaelli, Dunlop, and Fritsch 2013).

Having started at the top of the ladder of abstraction by arguing that narratives are universal and having provided subcategories of concepts that move down the ladder of abstraction that are useful to comparative research projects, we now turn to a more detailed discussion of how the NPF can specifically address the conceptual “traveling” problem in comparative public policy research.

4. TRAVELING WITH THE NPF

An example of the ongoing dilemma of the lack of universal conceptual categories in comparative policy studies was provided by Heidenheimer (1985) regarding the instability of Theodore Lowi’s now classic policy typology, which did not travel well when applied outside the US. This echoes Sartori’s lament on conceptual stretching and “misformation” that functions to undermine direct comparison between political systems. Indeed, Sartori is explicit in his prescription for development in comparative politics. “We do need, ultimately, ‘universal’ categories – concepts which are applicable to any time and place ... they must be *empirical* universals, that is, categories which somehow are amenable, in spite of their all-embracing ... abstract nature, to empirical testing” (Sartori 1970: 1035). While this plea went out almost 50 years ago, the need still exists in contemporary comparative policy studies, having been echoed in recent work by Atkinson and Coleman (1992), Peters (2013), Dodds (2013: 329–33), Henry et al. (2014: 299), and Wilder (2017).

So, how then does the NPF provide tools for this universal empiricism? To illustrate how the NPF’s structural categories (heroes, villains, victims) meet Sartori’s notion of a “high level universal category” we will contrast the NPF’s “hero” to Sartori’s discussion of the concept of “staff” in comparative public administration (Sartori 1970: 1042). Sartori notes that “staff” is more universal than “administration” since administration “retains some of the attributes associated with the more specific notion of ‘bureaucracy’” (Sartori 1970: 1042). An even less universal category to employ in comparative public administration is that of “civil service”, owing to its “associations with the modern State” (Sartori 1970: 1042). To arrive at the lowest level of abstraction in this conceptual category we might think of the particular attributes of bureaucrats in one nation compared to another. Thus, Sartori takes us down the ladder of abstraction, from the universal (staff) to the particular (national bureaucrat).

The NPF categories can mirror this fluid movement, as the “hero” (or other character) category is present in narratives across the world. The ability of the NPF to link heroes and their qualities (the hero who is a scientist, for example) can easily create subsets of heroes that are less than universal, but still have applicability in given contexts. Obviously, while all societies will have particular individuals prized for their wisdom and knowledge, the scientist-hero is predicated on the social acceptance of science as authoritative and the scientist as a trusted actor. This means that while this character may exist in numerous societies, there are at least certain segments of societies in which a scientist would not generally be seen as heroic: for instance, narratives produced by Young Earth Creationists in the US (McClure 2009).

Much the same approach can be taken to the other two core character categories – villains and victims – as can be done with the other major NPF elements (setting, plot, moral, and beliefs and strategies). Each is a universal and quantifiable structural category; however, they are also highly dependent on the qualitative cultural and political context in which the story originates. This flexibility allows for core NPF concepts to move up and down the ladder of abstraction while accommodating study-specific additions to core narrative elements.

As we attach more and more attributes to a moral or character category, for example, the concept becomes more concrete and anchored to a particular context, thus moving down Sartori’s ladder. Likewise, removing attributes moves us upward. This approach creates a link between thick description of qualitative studies and the generalizability of quantitative analysis afforded by NPF’s categories. By way of example of moving down the ladder of abstraction, qualitative and quantitative applications can utilize a network analysis-based “alter-ego” dyadic formulation for coding character types (see Weible et al. 2016). In the Weible et al. (2016) application of this approach their operationalization links heroes and/or villains with actions taken in the narrative, generating a distinction between heroic companies that create jobs and villainous companies that ravage the natural world. While companies may be either heroes or villains within policy subsystems across countries, the qualitative *reasoning* behind their inclusion in the category can vary. This approach is in accord with Sartori’s observation that “two items being compared must first belong to the same class, and either have or not have an attribute; and only if they have it, the two items can be matched in terms of which has it *more or less*” (Sartori 1970: 1038). This approach emerges in NPF studies which compare character categories in policy narratives from different nations (Lebel and Lebel 2018; O’Leary et al. 2017). To move back up the ladder of abstraction, one simply strips the categories of their more nuanced and contextual codes (in this case, heroic and villainous oil companies, which are not applicable to all policy contexts).

The NPF’s ontological and epistemological orientations allow for the empirical testing of hypotheses – related to universal narrative concepts – in a manner consistent with Sartori’s call. For example, the use of content analysis to quantify narrative components present in communications in both US and non-US settings (Huda 2018b; Weible et al. 2016) is complemented by the reliability of measures and findings at the micro level related to affect and support for particular character and morals advanced in a given policy narrative (Jones 2014).

While simple comparison of the number and type of characters is a common approach in many NPF studies (see Pierce, Smith-Walter, and Peterson 2014 for a review of NPF scholarship), other operationalizations can be undertaken using universal narrative elements. For example, Lebel and Lebel (2018) construct maps of ally networks between organizations and countries in the Mekong region. These networks were mapped to illustrate how documents that championed different problem solutions traveled between certain groups and nations. Lebel

and Lebel also found that actors strategically applied either the angel or devil shift strategy within their narratives to generate political mobilization (2018: 168). These strategic shifts have been seen in other policy areas studied by the NPF as well (Merry 2016; Shanahan et al. 2013; Smith-Walter et al. 2016).

Similarly, policy beliefs can be discerned from or within networks of actors. For example, Shanahan et al. (2008) constructed a distinction between preferences for state or national predominance in policy regarding Yellowstone National Park in the US by looking at the proportion of local and state heroes compared to federal heroes (called “allies”) (Shanahan et al. 2008). This is important to keep in mind, since the measure of beliefs associated with federalism emerged from utilizing universal conceptual categories associated with narratives that possessed a level of quantitative differentiation in their presence (see Sartori 1970: 1036–40). This means that characters and policy solutions can be used to build similar measures that travel across policy domains and national contexts.

As many comparative public policy studies have focused on institutional characteristics of governments and the impact of these variations on policy making (Adolino and Blake 2011; Lodge 2007), to the extent that the policies developed, adopted, and implemented can be understood to rely on the interplay between various forces to influence the agenda-setting process, the NPF can help to explore the differences and similarities between the rationale presented for (in)action. The NPF is also able to account for the varying influence that the bureaucracy, elected officials, political parties, and interest groups have in various policy domains by producing a general analysis of policy narratives put forth by both official and unofficial actors in the policy process (Weible et al. 2016).

The NPF maintains that while the content of policy narratives will vary across policy subsystems (Jones and McBeth 2010), narrative components will remain stable enough to travel across cultural contexts. Initial research applying the NPF to different subnational governments in the US and countries outside of the US seems to provide tentative support for this assumption (see Table 20.2). The next section will review the work of scholars who have applied the NPF in either non-US contexts and/or in an explicitly comparative manner.

4.1 Comparative NPF Applications

As of this writing few comparative policy studies have used the NPF. As such, a thorough research program has yet to be developed (Shanahan et al. 2017: 198–200). However, early comparative work can serve as “proof of concept”, that the NPF’s universality and multi-level conceptual compatibility can produce insights into differences and similarities in public policy development and implementation in various contexts.

For example, O’Bryan, Dunlop, and Radaelli (2014) conduct a comparative analysis of narratives about the “Arab Spring” in the UK House of Commons Select Committee on Foreign Affairs and the US House of Representatives Committee on Foreign Affairs. The study found that the expert testimony provided in committee hearings failed to counter the heuristics employed by elected officials, in either House (O’Bryan, Dunlop, and Radaelli 2014: 127). This suggests that despite structural differences in governing institutions, both failed to demonstrate that better substantive information about the “Arab Spring” led to learning by legislators.

Crow and Berggren (2014: 131) use a multi-case design to explore “narrative strategy, effectiveness, and framing of winners and losers” by stakeholders in Colorado’s environmen-

Table 20.2 *Listing of existing NPF studies with international or comparative focus*

NPF studies conducted on non-US countries		
Country	Topic	Author
Switzerland	Education policy	Schlauffer 2016
United Kingdom	Research use in policy decisions	Lawton and Rudd 2014
India	Municipal climate and air policy	Weible et al. 2016
India	Nuclear plant siting	Gupta, Ripberger, and Collins 2014
India	Agricultural biotechnology policy	Huda 2018a
Thailand	Education policy	Nakyam 2014
Korea	Construction permitting	Park 2014
EU	Impact assessments of the European Commission	Radaelli, Dunlop, and Fritsch 2013
Australia	Health policy	Fitzgerald 2013
Finland	Forest governance	Peltomaa, Hilden, and Huttunen 2016
Explicitly comparative NPF studies		
US/UK	Narratives of the Arab Spring in legislative committees	O'Bryan, Dunlop, and Radaelli 2014
Colorado	Four different environmental policies	Crow and Berggren 2014
Canada/EU/Australia/US	E-cigarette regulation	O'Leary et al. 2017
US County/US County	Local hydraulic fracturing policies	Gottlieb, Bertone, and Arnold 2018
Laos/Cambodia/Vietnam/Thailand	Water, energy, and food security	Lebel and Lebel 2018
South Africa/Norway	Climate change	Fløttum and Gjerstad 2016

tal policy arena. When media coverage and interest group communications were collected and analyzed, researchers found no difference in the use of science by either the winning or losing coalitions (Crow and Berggren 2014: 141). Additionally, winning narratives used villains more often than did losing narratives, though no differences between heroes and victim characters were identified and it was also discovered that winning narratives were associated with attribution of blame (Crow and Berggren 2014: 145).

O'Leary et al. (2017) are interested in exploring how claims made about vapor devices (e-cigarettes) in policy narratives are translated into regulations governing their use. They study policy narratives in documents produced by agencies in the European Union, Australia, the US, and Canada related to the risks and benefits associated with e-cigarettes as a part of the ongoing debate regarding their regulation. The researchers found that 60% of the claims regarding the threat posed by e-cigarettes were shared across the documents. Villains were always cast as the vapor devices and their manufacturers, and victims were reliably pictured to be “vapour device users, youth, and tobacco control” (O'Leary et al. 2017: 37).

Lebel and Lebel (2018) used Cultural Theory and the NPF to explore narratives surrounding the governance of water, energy, and food systems (nexus narratives) around the world and compare them to narratives in policy debates in the Mekong region of Southeast Asia. Using qualitative and quantitative approaches the researchers found significant differences between the manner in which three different active worldview categories (*Individualist*, *Hierarch*, and *Egalitarian*) chose to portray victims. They found policy narratives surrounding water, food, and energy security in the Mekong region were “significantly more likely to utilize social justice, and stakeholder dialogue themes in their narratives than non-Mekong texts” (Lebel and Lebel 2018: 169). It was also discovered that narratives from outside the Mekong region had been transmitted from international organizations.

Fløttum and Gjerstad marry the NPF with a “polyphonic” analysis of government reports to understand how authors combine selected voices to strategically orchestrate the construction

of the case for particular actions in their policy narratives. They find that the NPF offers an approach to the study of narratives that “has the potential to link individual narratives to political coalitions or worldviews” (Fløttum and Gjerstad 2016: 12).

Finally, a preliminary study of policy narratives regarding agricultural biotech policy in India focused on the ability of NPF categories to “travel”, by comparing English and Hindi media coverage (Huda 2018b). The key research questions address the nature of the variation in the use of specific narrative elements (characters, morals, and evidence) between English and Hindi media, and if Hindi language documents possessed particular features which differed from their English counterparts. The study found differences between English and Hindi-language publications in the manner in which bio-engineered (Bt) eggplant was described; the former explained how Bt eggplant kills pests, while the Hindi-language publication emphasized that Bt eggplant produces poison (Huda 2018b: 20–21). Hindi-language articles tended to contain fewer words and the English-language publication generated more articles. Importantly, however, while noting that the “results are exploratory in nature” (Huda 2018b: 14) Huda found no significant differences between the two sets of narratives in the frequency with which they deployed various narrative elements, lending initial support to the NPF assertion that the narrative elements appear to be generalizable across (at least certain) cultures and languages.

What should a comparative scholar take from the extant NPF scholarship? First, that the NPF can be used in comparative work to empirically compare policy narratives in different institutional contexts and generates non-obvious insights regarding similarities and differences *across* nations (Huda 2018a, 2018b; Lebel and Lebel 2018; O’Bryan, Dunlop, and Radaelli 2014; O’Leary et al. 2017). Second, that the NPF can be used to compare systems within nations and generate non-obvious insights regarding similarities and differences *within* nations (Crow and Berggren 2014; Gottlieb, Bertone, and Arnold, 2018). Finally, that the NPF’s categories can be employed universally and function as a scalable conceptual scaffolding that can link the micro-, meso-, and macro-levels of analysis within and across policy subsystems (Fløttum and Gjerstad 2016).

Another approach, of course, would be to compare policy narratives to see if particular manifestations of belief systems exist in similar policy areas across nations. Additionally, in areas where policy areas share cultural narratives, but produce different policy outcomes, factors other than narrative persuasion are likely to be more important in determining outcomes. This represents an important complement to existing institutional approaches, as narrative can serve as an “alternate hypothesis” that can systematically analyze and compare the role of political persuasion in policy.

The NPF might also complement comparative public policy interested in the impact of national myths on policy reform (Adolino and Blake 2011). Indeed, Berger (2009) suggests these myths provide guidance for contemporary approaches to constructing national identity as the stories “had heroes, who tragically died in defense of the nation ... and they also had villains, who contributed to the nation’s demise” (Berger 2009: 493). These origin stories require that scholars turn toward comparison with other nations to study whether they have an impact on present-day policy development.

5. LIMITATIONS

Of course, as with any approach to the study of policy the NPF has certain limitations and shortcomings. First, while narrative is universal to human societies, spaces where policy narratives can play out in public while reflecting genuine differences of opinion between groups are not. The NPF is likely to encounter difficulties when media environments and public records are not free and open.

Second, much NPF work has relied on deductive theories of belief systems from outside the framework to address the problem of narrative relativity. This means that while structural components of narratives are universal, the “identification of policy beliefs must be theoretically grounded” (Shanahan et al. 2017: 178). To achieve this grounding the NPF has drawn from Cultural Theory (Ney 2014), political ideology (Clemons, McBeth, and Kusko 2012) and others. However, it is vital these theories are valid and applicable to the various contexts they seek to compare. The universality of NPF structural components will not compensate for deductive theories which are not appropriately calibrated to the cultures in which they are applied.

Third, the NPF has traditionally relied on the relatively easy availability of public consumption documents on the internet, reports from public agencies, infrastructure that facilitates survey experimentation and the relative openness of public officials and members of civil society to free expression of policy preferences. These conditions may not hold in other contexts, and at least two NPF studies in India noted the difficulties associated with securing sufficient numbers of documents for meso-level content analysis (Huda 2018a; Weible et al. 2016).

Finally, the NPF is a young framework, and there are active conversations about how best to conceive of policy narratives (Weible et al. 2016). Whether there are more characters than the hero, villain, and victim, and even whether policy narratives can be shown to impact public policy outcomes are still open theoretical and empirical questions. As such, much work remains to be done, but early indications suggest that the study of narrative can offer much to the study of comparative public policy.

6. CONCLUSION

Public policy deals with issues of substantial import. Public policy makes statements about the values a society prizes, the behaviors it chooses to reward, and the actions it hastens to condemn. Human societies tell stories that promote particular ways of living, and these narratives exist in the most technical of bureaucratic reports, the most inspiring of presidential addresses, and at the tables of average citizens. In short, narratives are the way that humans make their experiences meaningful. The NPF can provide concepts that are universal, applicable at multiple levels of analysis, and empirically measurable. The shared human reliance on narrative means comparative public policy studies exploring contrasting stories about public action (or inaction) may be a path toward increased knowledge of the determinants of public policy in the disparate political systems around the world.

NOTE

1. While the NPF has developed its own model of individual cognition and associated narrative rationality, there does exist the possibility that this model is overly dependent on a Western-oriented approach to decision making. While we feel that we have laid out a strong case for the universality of narrative form, testing the universal nature of the *Homo narrans* model has not yet been undertaken (at least as an integrated totality). Should this model fail to be universally applicable, another formulation of narrative rationality that is more reliant on innate human faculties for assessing narratives is available for integration into the NPF. This model, the narrative paradigm, elaborated by Walter R. Fisher (1984, 1985, 1989), and extended by McClure (2009) relies on the innate ability of individuals to evaluate stories using two key criteria: narrative probability and narrative fidelity (Fisher 1989).

Narrative probability (as reformulated by McClure 2009) is composed of three separate assessments. The first, argumentative coherence is whether the structure of the narrative is one that is able to be arranged by the auditor in a meaningful manner. The second, material coherence, is assessed by the auditor when they compare this story to other stories on the same subject, and characterological coherence results from the auditor interrogating the “reliability of the character” (McClure 2009: 192). Narrative fidelity also has three components: whether the narrative comports with experiences the auditor has had, whether the narrative is commensurate with similar accounts of the issue they have previously found to be accurate, and whether the narrative presents attractive values and solid reasoning (McClure 2009: 192).

Each of these assessments relies on either the form and decipherability of the narrative itself or the lived experience of the individual performing the evaluation. Neither of these can be argued to exist as a fundamentally Western approach to evaluating narrative. Thus, even if the *Homo narrans* model does not adequately model decision making in non-Western settings, it is possible for comparative NPF studies to resort to a more general understanding of how narratives are evaluated for acceptability (for examples of the narrative paradigm applied in comparative and non-western contexts see Stroud 2003, 2004).

REFERENCES

- Adolino, J. R. and Blake, C. H. (2011). *Comparing Public Policies: Issues and Choices in Industrialized Countries*. Washington, DC: CQ Press.
- Ash, S., Moore, P., Vesely, L., and Grossman, M. (2007). The decline of narrative discourse in Alzheimer’s disease. *Brain and Language* 103: 181–2.
- Atkinson, M. M. and Coleman, W. D. (1992). Policy networks, policy communities and the problems of governance. *Governance: An International Journal of Policy and Administration* 5 (2): 154–80.
- Bennett, L. W. and Edelman, M. (1985). Toward a new political narrative. *Journal of Communication* 35 (4): 156–71.
- Berger, S. (2009). On the role of myths and history in the construction of national identity in modern Europe. *European History Quarterly* 39 (3): 490–502.
- Berinsky, A. J. and Kinder, D. R. (2006). Making sense of issues through media frames: understanding the Kosovo crisis. *Journal of Politics* 68 (3): 640–656.
- Blair, B. D. and McCormack, L. (2016). Applying the narrative policy framework to the issues surrounding hydraulic fracturing within the news media: a research note. *Research and Politics* 3 (1): 1–3.
- Booker, C. (2004). *The Seven Basic Plots: Why We Tell Stories*. New York: Bloomsbury.
- Clemons, R. S., McBeth, M. K., and Kusko, E. (2012). Understanding the role of policy narratives and the public policy arena: obesity as a lesson in public policy development. *World Medical and Health Policy* 4 (2): 1–26.
- Crow, D. A. and Berggren, J. (2014). Using the narrative policy framework to understand stakeholder strategy and effectiveness: a multi-case analysis. In M. D. Jones, E. A. Shanahan, and M. K. McBeth (Eds.), *The Science of Stories: Applications of the Narrative Policy Framework*. New York: Palgrave Macmillan, 131–56.

- DeLeon, P. (1997). *Democracy and the Policy Sciences*. Albany, NY: State University of New York Press.
- DeLeon, P. (2006). The historical roots of the field. In M. Moran, M. Rein, and R. E. Goodin (Eds.), *The Oxford Handbook of Public Policy*. New York: Oxford University Press, 36–58.
- Dodds, A. (2013). *Comparative Public Policy*. New York: Palgrave Macmillan.
- Dodge, J. (2015). Indication and inference: reflections on the challenge of mixing paradigms in the Narrative Policy Framework. *Critical Policy Studies* 9 (3): 361–7.
- Douglas, M. and Wildavsky, A. (1983). *Risk and Culture: An Essay on the Selection of Technological and Environmental Dangers*. Berkeley, CA: University of California Press.
- Edelman, M. (1985). *The Symbolic Uses of Politics*. Urbana, IL: University of Illinois Press.
- Ertas, N. (2015). Policy narratives and public opinion concerning charter schools. *Politics & Policy* 43 (3): 426–51.
- Fisher, W. R. (1984). Narration as a human communication paradigm: the case of public moral argument. *Communication Monographs* 51: 1–22.
- Fisher, W. R. (1985). The narrative paradigm: an elaboration. *Communication Monographs* 52: 347–67.
- Fisher, W. R. (1989). *Human Communication as Narration: Toward a Philosophy of Reason, Value, and Action*. Columbia, SC: University of South Carolina Press.
- Fitzgerald, J. L. (2013). Supervised injecting facilities: a case study of contrasting narratives in a contested health policy arena. *Critical Public Health* 23 (1): 77–94.
- Fløttum, K. and Gjerstad, Ø. (2016). Narratives in climate change discourse. *WIREs Climate Change* 8: e429. doi:10.1002/wcc.429.
- Goodin, R. E., Rein, M., and Moran, M. (2006). The public and its policies. In M. Moran, M. Rein, and R. E. Goodin (Eds.), *The Oxford Handbook of Public Policy*. Oxford: Oxford University Press, 3–36.
- Gottlieb, M., Bertone, E., and Arnold, G. (2018). “No fracking way” vs. “Drill baby drill”: a restructuring of who is pitted against whom in the Narrative Policy Framework. *Policy Studies Journal* 46 (4): 798–827.
- Gray, G. and Jones, M. D. (2016). A qualitative narrative policy framework? Examining the policy narratives of US campaign finance regulatory reform. *Public Policy and Administration* 31 (3): 193–220.
- Green, M. C. and Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology* 79 (5): 701–21.
- Gupta, K. (2012). Comparing public policy: using the comparative method to advance our understanding of the policy process. *Policy Studies Journal* 40 (1): 11–26.
- Gupta, K., Ripberger, J. T., and Collins, S. (2014). The strategic use of policy narratives: Jaitapur and the politics of siting a nuclear power plant in India. In M. D. Jones, E. A. Shanahan, and M. K. McBeth (Eds.), *The Science of Stories: Applications of the Narrative Policy Framework*. New York: Palgrave Macmillan, 89–106.
- Heidenheimer, A. J. (1985). Comparative public policy at the crossroads. *Journal of Public Policy* 5 (4): 441–65.
- Henry, A. D., Ingold, K., Nohrstedt, D., and Weible, C. M. (2014). Policy change in comparative contexts: applying the Advocacy Coalition Framework outside of Western Europe and North America. *Journal of Comparative Policy Analysis: Research and Practice* 16 (4): 299–312.
- Herman, D. (2009). *Basic Elements of Narrative*. Oxford: Wiley-Blackwell.
- Huda, J. (2018a). An examination of policy narratives in agricultural biotechnology policy in India. *World Affairs* 181 (1): 42–68.
- Huda, J. (2018b). Do language and culture matter in the policy process? Linguistic and cultural context in the Narrative Policy Framework. Presentation at the 76th Annual Midwest Political Science Association Conference, Chicago, IL. Cited with permission.
- Jones, M. D. (2014). Cultural characters and climate change: how heroes shape our perception of climate science. *Social Science Quarterly* 95 (1): 1–39.
- Jones, M. D., Fløttum, K., and Gjerstad, Ø. (2017). Stories about climate change: the influence of language on Norwegian public opinion. In K. Fløttum (Ed.), *The Role of Language in the Climate Change Debate*. New York: Routledge, 65–84.
- Jones, M. D. and McBeth, M. (2010). A narrative policy framework: clear enough to be wrong? *Policy Studies Journal* 38 (2): 329–53.

- Jones, M. D., McBeth, M. K., and Shanahan, E. A. (2014). Introducing the Narrative Policy Framework. In M. D. Jones, E. A. Shanahan, and M. K. McBeth (Eds.), *The Science of Stories: Applications of the Narrative Policy Framework*. New York: Palgrave Macmillan, 1–26.
- Jones, M. D. and Radaelli, C. M. (2015). The Narrative Policy Framework: child or monster? *Critical Policy Studies* 9 (3): 339–55.
- Jones, M. D. and Song, G. (2014). Making sense of climate change: how story frames shape cognition. *Political Psychology* 35 (4): 447–76.
- Jorgensen, P., Song, G., and Jones, M. D. (2018). Public support for campaign finance reform: the role of policy narratives, cultural predispositions, and political knowledge in collective policy preference formation. *Social Science Quarterly* 99 (1): 216–30.
- Kahan, D. M. and Braman, D. (2006). Cultural cognition and public policy. *Yale Law & Policy Review* 24: 147–70.
- Kahan, D. M., Braman, D., Gastil, J., Slovic, P., and Mertz, C. K. (2007). Culture and identity-protective cognition: explaining the white-male effect in risk perception. *Journal of Empirical Legal Studies* 4 (3): 465–505.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.
- Lawton, R. N. and Rudd, M. A. (2014). A narrative policy approach to environmental conservation. *AMBIO* 43: 849–57.
- Lebel, L. and Lebel, B. (2018). Nexus narratives and resource insecurities in the Mekong region. *Environmental Science & Policy* 90: 164–72.
- Lin, P.-Y., Sparks Grewal, N., Morin, C., Johnson, W. D., and Zak, P. J. (2013). Oxytocin increases the influence of public service advertisements. *PLoS One* 8 (2): e56934.
- Lodge, M. (2007). Comparative public policy. In F. Fischer, G. J. Miller, and M. S. Sidney (Eds.), *Handbook of Public Policy Analysis: Theory, Politics, Methods*. Boca Raton, FL: CRC Press, 273–88.
- Lodge, M. and Taber, C. S. (2005). The automaticity of affect for political leaders, groups, and issues: an experimental test of the hot cognition hypothesis. *Political Psychology* 26 (3): 455–82.
- Lodge, M. and Taber, C. S. (2007). The rationalizing voter: unconscious thought in political information processing. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1077972, accessed November 20, 2019.
- Lybecker, D. L., McBeth, M. K., and Stoutenborough, J. W. (2016). Do we understand what the public hears? Stakeholders' preferred communication choices for discussing river issues with the public. *Review of Policy Research* 33 (4): 376–92.
- Majone, G. (1989). *Evidence, Argument, and Persuasion in the Policy Process*. New Haven, CT: Yale University Press.
- Mar, R. A. (2004). The neuropsychology of narrative: story comprehension, story production and their interrelation. *Neuropsychologia* 42 (10): 1414–34.
- Mattilla, A. S. (2000). The role of narratives in the advertising of experiential services. *Journal of Service Research* 3 (1): 35–45.
- McBeth, M. K., Lybecker, D. L., and Garner, K. A. (2010). The story of good citizenship: framing public policy in the context of duty-based versus engaged citizenship. *Politics & Policy* 38 (1): 1–23.
- McBeth, M. K., Shanahan, E. A., and Jones, M. D. (2005). The science of storytelling: measuring policy beliefs in Greater Yellowstone. *Society and Natural Resources* 18 (May/June): 413–29.
- McBeth, M. K., Shanahan, E. A., Tigert, L. E., Hathaway, P. L., and Sampson, L. J. (2010). Buffalo tales: interest group policy stories in Greater Yellowstone. *Policy Sciences* 43 (4): 391–409.
- McClure, K. (2009). Resurrecting the narrative paradigm: identification and the case of young earth creationism. *Rhetoric Society Quarterly* 39 (2): 189–211.
- Merry, M. K. (2016). Constructing policy narratives in 140 characters or less: the case of gun policy organizations. *Policy Studies Journal* 44 (4): 373–95.
- Morris, J., Squires, N. K., Taber, C. S., and Lodge, M. (2003). Activation of political attitudes: a psychophysiological examination of the hot cognition hypothesis. *Political Psychology* 24 (4): 727–45.
- Nakyam, S. (2014). Educational decentralization policy in Thailand: unpacking its labyrinth to pinpoint an appropriately further step. Paper presented at International Conference on Public Administration, Chengdu Shi, China, October 24–26.

- Ney, S. (2014). The governance of social innovation: connecting meso and macro levels of analysis. In M. D. Jones, E. A. Shanahan, and M. K. McBeth (Eds.), *The Science of Stories: Applications of the Narrative Policy Framework*. New York: Palgrave Macmillan, 207–34.
- O’Byran, T., Dunlop, C. A., and Radaelli, C. M. (2014). Narrating the “Arab Spring”: where expertise meets heuristics in legislative hearings. In M. D. Jones, E. A. Shanahan, and M. K. McBeth (Eds.), *The Science of Stories: Applications of the Narrative Policy Framework*. New York: Palgrave Macmillan, 107–29.
- O’Leary, R., Borland, R., Stockwell, T., and MacDonald, M. (2017). Claims in vapour device (e-cigarette) regulation: a Narrative Policy Framework analysis. *International Journal of Drug Policy* 44: 31–40.
- Park, Y. S. (2014). A study of the construction permit process of 2nd Lotte World (skyscraper) using the Narrative Policy Framework. *The Korean Governance Review* 21 (2): 101–25.
- Peltomaa, J., Hilden, M., and Huttunen, S. (2016). Translating institutional change: forest journals as diverse policy actors. *Forest Policy and Economics* 70: 172–80.
- Peters, B. G. (2013). *Strategies for Comparative Research in Political Science*. New York: Palgrave Macmillan.
- Peterson, H. L. (2018). Political information has bright colors: narrative attention theory. *Policy Studies Journal* 46 (4): 828–42.
- Pierce, J. J., Smith-Walter, A., and Peterson, H. L. (2014). Research design and the narrative policy framework. In M. D. Jones, E. A. Shanahan, and M. K. McBeth (Eds.), *The Science of Stories: Applications of the Narrative Policy Framework in Public Policy Analysis*. New York: Palgrave Macmillan, 27–44.
- Polkinghorne, D. (1988). *Narrative Knowing and the Human Sciences*. Albany, NY: SUNY Press.
- Radaelli, C. M., Dunlop, C. A., and Fritsch, O. (2013). Narrating impact assessments in the European Union. *European Political Science* 12: 500–521.
- Riker, W. H. (1986). *The Art of Political Manipulation*. New Haven, CT: Yale University Press.
- Roe, E. (1994). *Narrative Policy Analysis: Theory and Practice*. Durham, NC: Duke University Press.
- Sabatier, P. A., Hunter, S., and McLaughlin, S. (1987). The devil shift: perceptions and misperceptions of opponents. *Western Political Quarterly* 41: 449–76.
- Sartori, G. (1970). Concept misformation in comparative politics. *American Political Science Review* 64 (4): 1033–53.
- Schattschneider, E. E. (1960). *The Semi-Sovereign People*. New York: Holt, Rinehart & Winston.
- Schlauffer, C. (2016). The narrative uses of evidence. *Policy Studies Journal* 46 (1): 90–118.
- Schmidt, V. A. (2010). Taking ideas and discourse seriously: explaining change through discursive institutionalism as the fourth “new institutionalism”. *European Political Science Review* 2 (1): 1–25.
- Shanahan, E. A., Jones, M. D., and McBeth, M. K. (2011). Policy narratives and policy processes. *Policy Studies Journal* 39 (3): 535–61.
- Shanahan, E. A., Jones, M. D., McBeth, M. K., and Lane, R. R. (2013). An angel on the wind: how heroic policy narratives shape policy realities. *Policy Studies Journal* 41 (3): 453–83.
- Shanahan, E. A., Jones, M. D., McBeth, M. K., and Radaelli, C. M. (2017). The Narrative Policy Framework. In C. M. Weible and P. A. Sabatier (Eds.), *Theories of the Policy Process* (4th edition). Boulder, CO: Westview Press, 173–214.
- Shanahan, E. A., McBeth, M. K., Hathaway, P. L., and Arnell, R. J. (2008). Conduit or contributor? The role of media in policy change theory. *Policy Sciences* 41: 115–38.
- Shanahan, E. A., Raile, E., French, K., and McEvoy, J. (2018). Bounded stories. *Policy Studies Journal* 46 (4): 922–48.
- Simon, H. A. (1947). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*. New York: Macmillan.
- Smith, D., Schlaepfer, P., Major, K., Dyble, M., Page, A. E., Thompson, J., Chaudhary, N., Salali, G. D., Mace, R., Astete, L., Ngales, M., Vinicius, L., and Bamberg Migliano, A. (2017). Cooperation and the evolution of hunter-gatherer storytelling. *Nature Communications* 8: 1–9.
- Smith-Walter, A. (2018). Victims of health care reform: Hirschman’s rhetoric of reaction in the shadow of federalism. *Policy Studies Journal* 46 (4): 894–921.
- Smith-Walter, A., Peterson, H. L., Jones, M. D., and Reynolds, A. (2016). Gun stories: how evidence shapes firearms policy in the US. *Politics & Policy*. doi:10.1111/polp.12187.
- Stone, D. (2012). *Policy Paradox: The Art of Political Decision Making*. New York: W. W. Norton.

- Stroud, N. J. (2008). Media use and political predispositions: revisiting the concept of selective exposure. *Political Behavior* 30: 341–66.
- Stroud, S. R. (2003). Narrative translation across cultures: from the Bhagavad Gita to the Legend of Bagger Vance. *Journal of Communication and Religion* 26 (1): 51–82.
- Stroud, S. R. (2004). Narrative as argument in Indian philosophy: the Astavakra Gita as multivalent narrative. *Philosophy & Rhetoric* 37 (1): 42–71.
- Sugiyama, M. S. (2001). Food, foragers, and folklore: the role of narrative in human subsistence. *Evolution and Human Behavior* 22: 221–40.
- Taber, C. S. and Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science* 50 (3): 755–69.
- Thompson, M., Ellis, R., and Wildavsky, A. (1990). *Cultural Theory*. Boulder, CO: Westview Press.
- Weible, C. M., Olofsson, K. L., Costie, D. P., Katz, J. M., and Heikkila, T. (2016). Enhancing precision and clarity in the study of policy narratives: an analysis of climate and air issues in Delhi, India. *Review of Policy Research* 33 (4): 420–441.
- Wilder, M. (2017). Comparative public policy: origins, themes, new directions. *Policy Studies Journal* 45 (S1): S47–S66.
- Wong, J. (2014). Comparing beyond Europe and North America. In I. Engeli and C. Rothmayr Allison (Eds.), *Comparative Policy Studies: Conceptual and Methodological Challenges*. Basingstoke: Palgrave Macmillan, 163–94.
- Zak, P. J. (2015). Why inspiring stories make us react: the neuroscience of narrative. *Cerebrum: The Dana Forum on Brain Science* (February): 1–13. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4445577/pdf/cer-02-15.pdf>.
- Zak, P. J., Kurzban, R., and Matzner, W. T. (2005). Oxytocin is associated with human trustworthiness. *Hormones and Behavior* 48: 522–7.

PART VI

ISSUES FOR FURTHER RESEARCH

21. Trends in the development of comparative policy analysis

Iris Geva-May, David C. Hoffman and Joselyn Muhleisen

1. INTRODUCTION

In her 2013 article, “Policy Analysis Reaches Midlife”, and her seminal second book titled *Beyond Machiavelli: Policy Analysis Reaches Midlife* (2013), Beryl Radin highlights the *Journal of Comparative Policy Analysis: Research and Practice (JCPA)* and its mission statement as important stepping-stones in the development of the field of comparative policy analysis. Thomson Reuters’s Who’s Who (2008) states that its founder, Iris Geva-May, pioneered the domain. To date, *JCPA* is still the only journal exclusively devoted to comparative policy studies. Founded 20 years ago, the mission of *JCPA* has been advanced with the support of leading scholars and dedicated editorial board members, such as Laurence E. Lynn Jr., first co-editor, Peter deLeon, Duncan MacRae, David Weimer, Beryl Radin, Frans Van Nispen, Yukio Adachi, Claudia Scott, Allan Maslove and others in the US, Canada, Europe, Australia/New Zealand and Asia. In recent years it has increased from three to five issues per annual volume due to the number of high-quality submissions. There are about 120 articles submitted each year. In 2016 the *JCPA* acceptance rate was 21%, there were 443 citations of its articles, and the impact factor increased by 83% in 2018 from 1.07 to 1.862 (Thomson Reuters SSI Citation Index, June 2018). This indicates that an increasing number of policy researchers and scholars from many interdisciplinary policy domains, internationally, are engaging in comparative studies. An increasing number of new comparative policy books are being published and more courses in comparative public policy are being offered around the world. This upward trajectory in comparative policy studies is also evident in the establishment of the International Comparative Policy Analysis-Forum (ICPA-Forum) with over 1,400 international members, around 50 international institutions, and 12 partner associations involved together in a scholarly network advancing comparative studies.

In an effort to take stock of where the field of comparative policy analysis has been and where it is going, this chapter focuses on the results of a broad content analysis of comparative public policy articles indexed on the EBSCO Academic Complete database and in *JCPA*, the field’s flagship journal. This analysis categorizes redundant research articles published over the last two decades in comparative policy analytic studies by topic, methodology, and countries studied. This chapter begins with a brief survey of the history of the field, moves on to a discussion of its scope and importance, reviews the results of the content analysis, and finally makes suggestions for some new directions that comparative policy analysis in general, and *JCPA* in particular, might head in the future.

2. COMPARATIVE POLICY ANALYSIS DEVELOPMENT AS A DISTINCT FIELD OF STUDY

In the 1990s, comparative policy analysis and comparative policy analytic studies, as a modern research tradition, joined comparative politics and comparative public administration as a sub-domain of political science and public administration that uses the comparative method. Policy studies and policy analysis, deriving from political science and public administration, became new areas of social scientific inquiry only in the 1960s. The comparative layer, for all three domains, was a later addition. The field of comparative policy analysis and policy analytic studies developed from and drew on research traditions in comparative politics based on political science and public administration.

The rationale for the emergent field of policy analysis in the 1960s was the need for democratic societies and their elected policy makers to be transparent, efficient, and accountable. In the words of Wildavsky (1979), a pioneer of the policy analysis field, policy analysis is a craft that “speaks truth to power”. Initially, the policy analytic movement started with the notion that the key objective of public service was to design, enact, and implement better public policies – meeting the Kaldor-Hicks criteria for economic efficiency as Lindblom (1958) first asserted. Thus, a major cornerstone in policy analysis has been the principle that decision making should be systematic, evidence-based, verifiable, and evaluable (therefore, transparent and accountable) to meet democratic principles, social and economic needs, and be answerable to the public (Brans, Geva-May, and Howlett 2017). Evidence-based policy making also implies, by definition, looking for evidence “elsewhere” for historical, international, disciplinary, or other comparisons of data, facts, and events. This explicitly implies that the comparative perspective, at the micro- or macro-levels, is paramount for effective and efficient policy development. We witness a first, lone attempt at including “comparative public policy” as perceived today, in Dierkes, Weiler, and Antal’s 1987 book, *Comparative Policy Research: Learning from Experience*, in which Wildavsky is a contributor. Later Richard Rose published scholarship on the notion of “lesson drawing” as a comparative method (1993).

While comparative politics has been a domain of study since the late 1800s, its revival in the 1970s and 1980s has had a significant impact on the interrelated field of public administration and public policy. John Stuart Mill’s seminal *A System of Logic*, first published in 1843, is widely considered to be the first systematic formulation of the modern comparative method for social science research. The comparative method began to be systematically applied in the modern academic study of politics through discussions about scientific comparative research design and methodologies in the 1970s and 1980s. The movement can be said to have begun in 1968, when the journal *Comparative Politics* was established by the Graduate School at the University Center at City University of New York. Harold Lasswell published his “The Future of the Comparative Method” as the lead article of the first issue of this journal.

It is interesting to note that while comparative policy analytic studies shared a developmental trajectory with both comparative politics and comparative public administration, as Beryl Radin argues (2013a, 2013b), it has tended to thrive as comparative public administration stalled. Comparative public administration uses the comparative method to study bureaucratic institutions and identify policies that are applicable in different political contexts. Substantively, public administration focuses almost exclusively on the public bureaucracies of the executive branch and the implementation stage of policy, drawing a distinction between administration and politics.

After World War II, scholars began to use comparative studies to answer questions about development and find administrative strategies that would work in different countries. In 1960 the comparative administration group was created and its Chairman, Fred W. Riggs, used it to push forward a scholarly agenda for the field. Comparative public administration experienced a boom in the 1960s when several journals were formed to advance the field. However, in the 1970s funding and academic support ended and the discipline floundered (Heady 2001). Reflecting this decline in interest, *The Journal of Comparative Administration* was renamed *Administration & Society* in 1973. In the late 1980s and 1990s neoliberal economic principles changed the approach to government in the United Kingdom and United States. Instead of privatizing public authorities and services, governments embraced mechanisms that provide competition, applied market principles, and focused on efficiency and rightsizing.

International development organizations projected neoliberalism globally as part of the Washington Consensus. The new ideas about government were embodied in the New Public Management literature and this scholarship led to a slight resurgence of comparative public administration in the 1990s. Since that period, the discipline of comparative public administration has been stagnant. According to Jreisat, many scholars avoid cross-cultural studies or any study in which the units of comparison are not functional equivalents (2002). This is a significant shortcoming of comparative public administration as the interdependence that results from globalization necessitates such comparisons. As comparative policy analysis is not limited to focusing on the unit of bureaucratic administration, it has taken up the challenge of analyzing how policies work in similar and dissimilar environments, but it has also taken on the study policy process, design, and implementation that were previously in the domain of public administration.

The rise in output in the comparative policy analysis field and the concurrent decline of output in comparative administration are borne out by data gathered from the EBSCO Academic Complete database and from Google Books. Figures 21.1 and 21.2 show the growth of literature on Comparative Public Policy and the decline of literature in Comparative Public Administration. Figure 21.1 is a comparison of the number of articles indexed in the

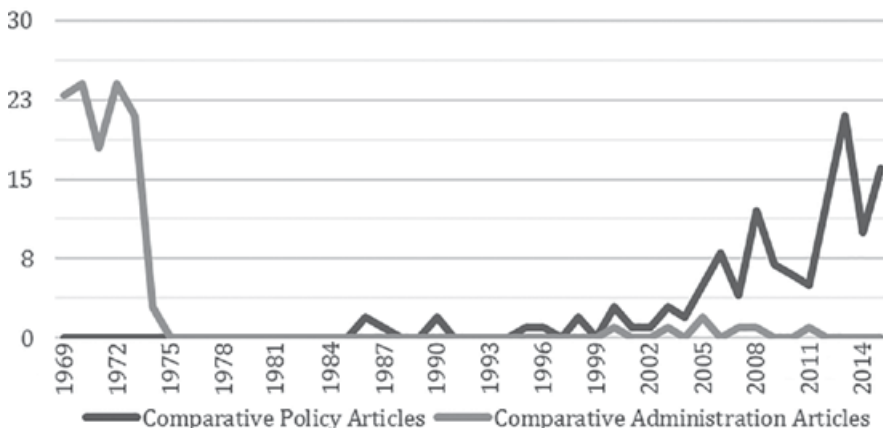


Figure 21.1 Comparative policy vs. comparative administration articles indexed in the EBSCO database by year



Note: The y-axis represents the percentage of two-word phrases in the corpus comprised by the target phrase. Although these two phrases will certainly not capture all the references to these fields, they are essentially a random sample of all references which can be expected to capture about the same portion of the total over time.

Figure 21.2 Relative usage rates of “comparative policy” and “comparative administration” by year in the “Google Million” database

EBSCO Academic Complete database per year with the words “comparative administration” and “comparative policy” in their titles or abstracts from 1969 to 2015. Figure 21.2 shows the result of a Google Books N-gram search for “comparative policy” and “comparative administration.” The chart represents the percentage of two-word combinations in the “Google Million”, a randomly sampled set of a million books, comprised by the target phrases.¹ Figure 21.3 in another Google Books N-gram search shows how much more frequent the phrase “comparative politics” is than either “comparative policy” or “comparative administration”. Of course there is ample comparative work that may not be explicitly labeled comparative. Unfortunately, our search strategy does not capture those studies that perform comparative work without using the term “comparative.” Although we acknowledge that this search method does not capture all the published literature which has not been explicitly termed “comparative”, we believe it shows strong trends in a fair sampling of overall publication over 20 years. The graphs look different because they show data from different samples and measure the prevalence of “comparative policy,” in different samples, but the overall trend that they show is the same: there is more discourse about comparative policy now than ever before. We believe that the fact that the same trend is apparent in different samples measured by different methods adds robustness to the finding.

3. THE IMPORTANCE AND LIMITATIONS OF COMPARATIVE POLICY ANALYSIS AND COMPARATIVE POLICY ANALYTIC STUDIES

Why do we compare public policies? We do so to be more effective and efficient, avoid the replication of failures, to maximize our use of resources, to save time, and to be inspired by those similarities that allow for degrees of lesson drawing. Comparative policy analysis and policy analytic studies contribute both to the tenet of evidence-based policy making and to that of efficiency and effectiveness, that is saving precious time and resources in an era when challenges need to be faced at optimal speed and level. It holds the promise of systematic policy making which is a cornerstone in modern democracies (Geva-May and Wildavsky 1997; Neustadt and May 1986; Ostrom 1998; Rose 1991; Rose and Mackenzie 1991; Weimer 1993; Weimer and Vining 1999; Wildavsky 1979).

It is natural to pursue comparisons within the field of policy analysis for many reasons. De facto comparisons, implicit or explicit, have always pervaded the work of social scientists. The basic research methods advocated in experimental designs are comparative in that they use control and experiment groups. To cite Swanson (1971: 145) “Thinking without comparisons is unthinkable and in the absence of comparisons, so are all scientific thought and scientific research.” Comparison is an innate cognitive attribute of humans that we apply in our daily lives, in our professional activities, or as producers or consumers of products. Public policy can be viewed as one such product. Further, there are political, cultural, sociological, economic, and other terms of reference within different contexts. What may be considered successful in one country may be assessed harshly in another due to different values as social controls.

But, context, and transfer/borrowing/lesson drawing *can be* compatible between social units: while transferring, borrowing, or lesson drawing are determined with a view of the country’s particular structure, culture, and politics (Geva-May 2002a, 2002b), in many cases “the commonalities are more important than the differences” (Ingraham 1996: 4). The value of

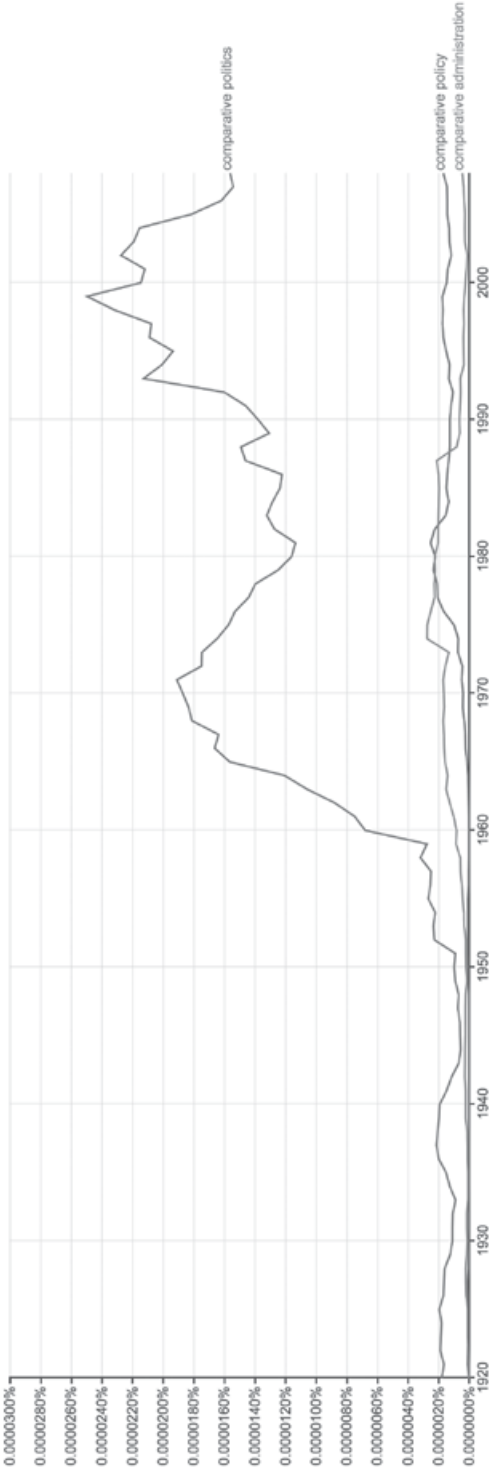


Figure 21.3 Relative usage rates of “comparative politics”, “comparative policy”, and “comparative administration” by year in the “Google Million” database

comparative policy studies lies in “crossing national boundaries [and] expand[ing] the number of programs that can be observed in action ... the fact that they are foreign introduces an element of speculation whether they can transfer. But speculation is bounded, for experience elsewhere provides palpable evidence how programs actually work” (Rose 1993: 110).

Comparisons are inherent in today’s public policy analytic world for at least four reasons. First, access: in the global village age, distance has been made less significant through communication technology and transportation systems. Public policies are more visible as are their successes or failures, which can be adopted or avoided, as would any tangible product. Second, economic path dependence or interdependence among national actors because of economic considerations: markets, customers, firms, legal systems and transactions affect the welfare or even survival of other entities. Third, in the global village governments are faced with similar national policy problems that usually have similar triggers and outcomes and lead to similar socio-economic, structural, and technological challenges and dilemmas. Loss or gain of human capital, population aging and skilled workforce scarcity, safety and security, healthcare access and so on, are only a few examples of such policy problems. Finally, cross-national problems constantly emerge and reveal the extent of interdependence (Geva-May 2002a, 2002b, 2004, 2008, 2015). Take for instance cross-border legal or illegal migration from one country to another affecting the welfare or socio-political canvass of the target state; or environmental policies: whether or not countries adhere to the Paris Agreement, their policies affect other countries’ pollution levels – air simply cannot be stopped at a jurisdiction’s borders. As long as there are governments and public services or institutions faced with similar emerging global issues, there are, as Rose (1991) contends, “lessons to be drawn”.

The “Aims and Scope” statement of *JCPA* embodies, to a large extent, the dimensions of the field of comparative policy analysis. These are: (1) Drawing lessons based on circumstances in which compared policy-related issues have in common certain manipulable policy, program, or institutional variables; (2) Contribution to comparative theory development; (3) Presentation of theory-based empirical research; (4) Comparative evaluations of research methods; (5) Comparative practice implications of theory-based research; and (6) Uses of conceptual heuristics to interpret practice (Geva-May and Lynn 1998).

Comparative methodology does have its limitations in public policy. The methods first explicated in Mill’s *A System of Logic* are difficult to apply in the social sciences in general and in comparative public policy analysis in particular, because it is problematic to find sufficiently similar cases to compare, and because of resources, timeliness, and other practical constraints. This leads to problems of randomization versus purposive limitations. Gerring addresses these limitations in his work on case study selection (Gerring 2006; Geva-May 2002a, 2002b). Furthermore, experimenting with policies over a *populus* that can be irreparably affected, deems the experimental/control method inapplicable or unethical.

Public policy comparisons also face the problem that policies or governmental structures – for instance in education, immigration, and transportation – are typical of the cultural or structural context in which they emerge. In Lasswell and Kaplan’s view (1950), “context is all-important.” Additionally, we need to acknowledge that public policy does not occur in laboratory conditions, and that there are limitations attached to the difficulty to observe and rigorously compare these “contexts” as control groups or experimental groups for ethical and practical reasons; the target units of comparison are purposive rather than random, and they are affected by volatile events, agendas and timelines, unexpected natural or economic developments, political changes, and so on. While in social science research we acknowledge external

interfering variables, the volatility of public policy implies significant interference of external and unexpected variables. What we are studying in comparative public policy analytic studies are not control and experimental groups but rather “naturally occurring experiments”; that is, what happens in a certain “social unit,” and what lessons can be drawn from one jurisdiction to others.

Recognizing the limitations we have just discussed, we need to define what comparative policy analytic studies can enhance given contextual differences. Primarily, in comparing and lesson drawing there are nuances on the comparative spectrum depending on the respective culture or system of said social units. They range from full Transfer, Borrowing, Adoption, to Adaptation, or mere “Pinching” (deLeon and Resnick-Terry 1998; Geva-May 2002a, 2002b) or in Rose’s (1991, 1993) words from Copying, Emulation, Hybridization, to mere Synthesis or Inspiration. This view of the nuanced nature of comparative public policy accounts for the perception that there must be core commonalities among aspects of policy problems, which are more similar than different and that they can be taken up at distinct levels. To address the validity of the comparison, one needs to identify what are the comparable common core aspects, and what are the similar or dissimilar causality factors, which dictate the search for comparative information. The contextual differences would then shape the degree to which lessons are drawn ranging from adoption to sparking a creative signal, which can lead to the creation of a policy solution that otherwise would not have been reached.

4. TRENDS IN COMPARATIVE POLICY ANALYSIS REFEREED ARTICLES: CONTENT ANALYSIS OF EBSCO ALL ACADEMIC DATABASE AND THE JCPA 1998–2016

In order to track the growth of the field of comparative policy analytic scholarship in the last two decades and discover any prevalent trends within the domain, we conducted a content analysis of 490 articles in two refereed sets: (a) 144 articles contained within the EBSCO Academic Search Complete database, identified as “comparative” and published between 1976 and 2016. They were returned when the phases “comparative policy analysis” or “comparative public policy” were entered; (b) 356 articles that have appeared in *JCPA* throughout 1998 to 2016, and which were all comparative by definition as they were published in line with this journal’s specific focus. The former set of articles was not inclusive of the *JCPA* articles, and this is another reason for deciding to use this data set.

Although there is undoubtedly more literature that is substantially about comparative policy that was captured by our search (for instance, a Google Scholar search for “comparative policy” yields more than 18,000 hits), our search provides a sample of the overall literature of a manageable size that provides a window on the total output. The EBSCO database was chosen for these practical reasons, although any broad database of academic articles might work equally well to provide a sample of academic output that is essentially random. For both these sets of articles, we used the program NVIVO to code each article for (1) Year of Publication, (2) Policy Area, (3) Country of the Authors’ Institutions, (4) Number of Countries Studied, (5) Names of Countries Studied, (6) Journal of Publication, and (7) Methodology and/or Theory. These categories were chosen because they could be detected by reading the abstracts of the articles in most cases, and because they could provide useful information about the type of work that is, or is not, being produced in the comparative policy field.

Not all the data collected figure into the present analysis. Here we concentrate on the results concerning year published, policy area, countries studied, and methodology/theory. The categories under the headings of Policy Areas Studied and Methodology Employed were arrived at inductively. They certainly do not represent the only ways of classifying articles. All coding judgments have been reviewed by at least two coders, and none by more than three. All discrepancies were resolved in conference between coders. We therefore believe in the robustness, reliability, and validity of the gathered evidence, and that it strongly supports a number of important claims about the growth of the field of comparative policy analysis, and about the growth of comparative public policy analytic scholarship in the last two decades.

4.1 The Growth of the Field of Comparative Policy Analysis and Policy Analytic Studies

There are strong indications that general interest in the field of comparative policy analysis is increasing. As shown by Figure 21.1, even excluding *JCPA*, which is not indexed by EBSCO, the number of journal articles published whose titles and abstracts contain some combination of the terms “comparative policy” or “comparative public policy” has increased dramatically over the course of the last 15 years. These results should be viewed in light of the fact that the overall number of academic journals indexed in EBSCO has likely increased as a whole, a fact that needs to be controlled for in future analysis. Also, it should be noted that, like the *JCPA*, there may be other journals that are not included in the EBSCO database.

While the overall number of publications in the field of public policy analysis has been on the rise, *JCPA* remains central to the field as the only journal devoted solely to comparative policy analysis and policy analytic studies, with very few refereed journals in our EBSCO sample having published more than a couple of comparative articles.

Next, we turn to a more detailed analysis of what has been published in the journals archived in the EBSCO database and the *JCPA*.

4.2 Comparative Policy Analysis Articles in the EBSCO Academic Complete Database: Publication Trends

A search of the EBSCO Academic Complete database for articles in comparative policy analysis yielded 144 hits for the search period 1976 to 2016, exclusive of review pieces. The first hit was in 1986. Of the 88 journals in the sample that had published at least one explicit article in comparative policy analysis between 1976 and 2016, only four had published five or more articles, while the vast majority, 75 journals, had published only one. One journal that stands out is *The Journal of European Public Policy and Politics & Policy*, which had ten comparative articles. Also competitive were the *Journal of Policy & Practice in Intellectual Disabilities* (8 articles), *Policy Studies* (5 articles) and *Comparative Political Studies* and *Social Policy and Administration* with 4 articles each.

Although the date range of the search is broader, two decades, than the publication period of *JCPA* since its founding in 1998, we believe that this provides a robust comparison to the field of comparative public policy analytic studies as it exists beyond the *JCPA*.

Table 21.1 *EBSCO articles published in common policy areas, 1976–2016*

Policy area studied	Percentage	Articles
Health, Healthcare, and Medicine	24%	34
Markets, Money, and the Economy	19%	28
Welfare, Social Services, and Poverty	15%	22
Local and Regional Government & Policy	13%	18
Education	10%	15
International Relations and Organizations	9%	13
Science & Technology	8%	12
Environment	7%	10
Human Rights	6%	8
Regulation & Deregulation	4%	6
Media & Communication	3%	5
Immigration and Migration	3%	4
Globalization	2%	3
Privatization and Public–Private Partnerships	2%	3
Criminal Justice	1%	2
Development of Underdeveloped Regions	1%	2
NGOs and Non-Profits	1%	2

Note: Policy areas are not exclusive categories. That is, each of the total 144 articles coded from the EBSCO database may have been coded as studying multiple policy areas.

4.2.1 Policy areas studied

In general, in the matter of policy areas studied, the top four topics in the EBSCO analysis are the same as it will be observed in the *JCPA*, although in a slightly different order. There is relatively more attention to health and healthcare in the EBSCO sample (24%) than in *JCPA* (11%) (see *JCPA* Content Analysis below). Both the EBSCO samples and the *JCPA* lack much coverage in the areas of criminal justice, human rights, arts and culture, and developmental aid.

Note that policy areas were coded in a non-mutually exclusive way to allow for one article to count in two or more areas. In addition, 21% of articles had one or more areas not captured in the list and were categorized as “other”. Many of these articles were devoted to building theory or frameworks (Table 21.1).

4.2.2 Countries studied

The results of the EBSCO analysis also look very similar to those of the *JCPA* analysis in terms of the countries that have been studied. North America and Europe predominate, and Central America, Africa, the Middle East, and Central Asia are the least studied regions. A number of new countries – like Iran, Haiti, and Cuba – do show up in the EBSCO articles, but these are mostly from comparative policy briefs published in a single issue of one journal, the *Journal of Policy & Practice in Intellectual Disabilities*, in 2008 (Figure 21.4).

4.2.3 Methodology and theory

The results of the analysis of methodology and number of countries studied show much the same tendencies for the EBSCO and *JCPA* samples. The largest number of articles in the EBSCO data set involved historical policy comparisons or were case studies (53%). Theory and Framework pieces follow at 42% of the comparative articles identified. Although these methods are still ranked toward the bottom of the list of all articles in the EBSCO sample, the



Figure 21.4 Articles per country in the EBSCO sample

8% engaging in comparative Ethnographic and Interview techniques were more frequent than in *JCPA* (2%).

Methodology was assessed in a way that does not involve mutual exclusion: a single article could be fit into more than one category as was appropriate. The following are descriptions of our methodological categories. *Case-Based and Historical*: Articles in this category draw conclusions by comparing different countries, localities, and policy regimes, often over a period of time. They do not attempt to delimit any set of texts or sources for analysis, but rather freely mix scholarship, journalism, government reports, and other sources. *Theory and Framework Pieces*: Articles in this category have the advancement of a theoretical or methodological framework as one of their aims, as opposed to providing insight into specific cases or answering an empirical question. Frequently they advance this theoretical or methodological objective by employing some other more specific method. For instance, theoretical perspectives are often advanced through comparative case studies. *Quantitative Analysis*: Articles in this category employ descriptive statistics, significance testing, regression and cluster analyses, and other statistical methods. *Textual and Content Analysis and Qualitative Research*: Articles in this category delimit a set of texts – such as speeches, newspaper articles, or legislation – for specific analysis, and then employ methods like content and rhetorical analysis to draw conclusions. *Survey Created for the Specific Research*: The authors of articles in this category created and conducted surveys specifically for the articles. *Ethnography and Interview*: Articles in this category employ ethnographic fieldwork and in-depth interview to draw conclusions (Table 21.2).

Table 21.2 EBSCO methodology, 1976–2016

Methodology employed	Percentage	Article
Historical and Case-Based	53%	76
Theory and Framework Pieces	42%	60
Quantitative Analysis	19%	28
Textual, Content & Qualitative Analysis	11%	16
Ethnography and Interview	8%	11
Survey Created for Specific Research	3%	5

5. THE JOURNAL OF COMPARATIVE POLICY ANALYSIS

5.1 Publication Trends

Our content analysis revealed a number of distinct publishing trends in the *JCPA*. As aforementioned, it is interesting to note the high consistency of trends: the general tendencies are almost the same in the *JCPA* (devoted exclusively to comparative policy scholarship) as in other journals presenting comparative policy studies, in terms of the areas studied, countries, and topics.

5.1.1 Policy areas studied

JCPA has a very strong history of publication in the area of economic and financial policy, on the one hand, and of welfare states and social subsidies, on the other. There has also been a strong interest in international agreements and organizations, the environment, education, healthcare, and immigration. Comparatively, as in the EBSCO identified comparative policy publications, there is little that has been published about criminal justice, economic development, foreign policy, human rights, NGOs, and arts and culture. At first glance it would appear the *JCPA* has not much partaken in the trend of publishing centered on the key term “globalization”. However, such a view would not take into account frequent publication on a number of overlapping topics, including “convergence” and “Europeanization” (Table 21.3).

Table 21.3 *JCPA* articles published in common policy areas, 1998–2016

Policy areas studied	Percentage	Articles
Markets, Money, and the Economy	19%	68
Welfare, Social Services, and Poverty	14%	49
Health, Healthcare, and Medicine	11%	40
Local and Regional Government & Policy	11%	38
International Relations and Organizations	10%	37
Environment	9%	32
Education	9%	31
Media & Communication	6%	20
Science & Technology	6%	20
Immigration and Migration	5%	19
Regulation & Deregulation	5%	18
Privatization and Public–Private Partnerships	4%	16
Disasters	3%	9
Globalization	3%	9
Development of Underdeveloped Regions	2%	7
Human Rights	2%	6
NGOs and Non-Profits	1%	5

5.1.2 Countries studied

While *JCPA* has published work about countries on all five continents, studies have tended to focus on the most economically developed countries. The United States is the single most-studied country, with 52 articles including it in comparisons, and Europe is the most studied region. Central America, the Caribbean, Africa, the Middle East, and Central Asia have

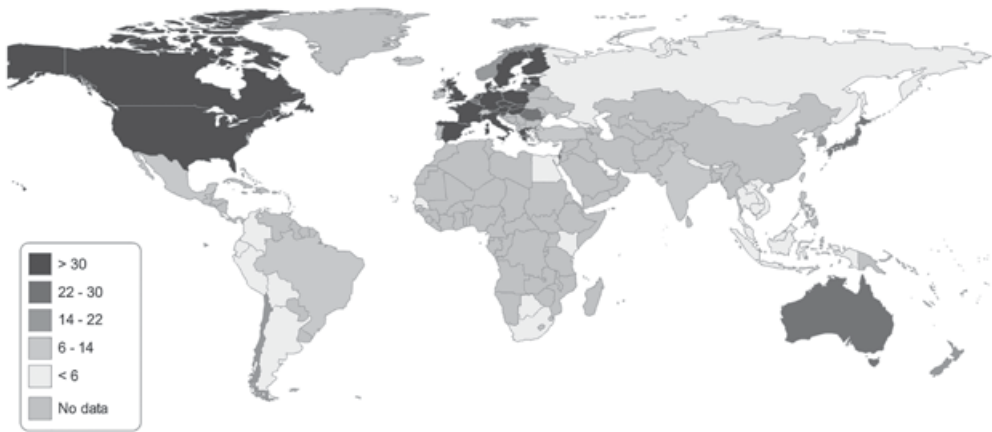


Figure 21.5 Articles per country in JCPA

received relatively sparse coverage. Two international organizations frequently determined the set of countries compared, the European Union (EU) (21 articles, 6%) and the Organisation for Economic Co-operation and Development (OECD) (9 articles, 3%). Countries named in EU or OECD studies are not included in the totals used for Figure 21.5 because it was frequently difficult to determine the sub-set of member states that were being studied.

The following is a breakdown of the methodologies employed in the *JCPA* articles. The single largest methodological category was the case study, which represented 62% of the total sample. Relatively few articles (2%) drew upon interview or ethnographic methods. A sizable portion (42%) of the articles attempted to employ or develop a theoretical or methodological framework, but a greater portion (58%) of articles compared practices and outcomes without advancing a theoretical framework (Table 21.4).

Table 21.4 *JCPA* article methodology, 1998–2016

Methodology employed	Percentage	Articles
Case-Based & Historical	62%	219
Theory and Framework Pieces	42%	150
Quantitative Analysis	15%	55
Textual & Content Analysis & Qualitative	10%	34
Survey Created for this Research	3%	11
Ethnography and Interview	2%	7

6. CONCLUSIONS

The content analysis study presented in this chapter has sought to identify the development and trajectory of comparative policy studies. The analysis revealed that the strengths and weaknesses of the comparative policy analysis in the field at large in terms of topics studied and methods employed, are similar to the strengths and weaknesses of the specifically oriented lead journal in the field, the *JCPA*.

The domain of comparative policy analysis and public policy analytic studies has gained momentum, as Figures 21.1 and 21.2 show, after 1998 as the field of comparative administration declined. The new comparative domain derived from and extended what the “old” social sciences research did not: robust, actionable findings grounded in comparison to decision makers and to the scholarly community. Since 1998, it has followed an abrupt upward trajectory, both in terms of scholarship as well as of regional coverage. *JCPA* has played an important role in this development by channeling a multitude of policy disciplines, regional cases, and methodological and theoretical approaches. It is still the only journal of comparative public policy devoted entirely to the publication of articles on comparative policy analytic studies. Although we cannot know exactly the publication output of the entire comparative policy analysis field, *JCPA*’s output accounts for a sizable portion of it and is an indicator of the interest in and trend of comparative policy studies.

In the area of methodology, while the majority of articles published are still case-based comparisons, innovative theoretical frameworks increasingly contribute to the comparative knowledge base. We note in the *JCPA*, theoretical dialogues on dependence, multiple streams (as in the highly cited Béland and Howlett 2016), and convergence (Clavier 2010; Coleman 2001; Happaerts and Van Den Brande 2011; Kuipers 2009) have taken place. Notwithstanding, comparative development of thought in other prominent frameworks such as punctuated equilibrium theory, and policy diffusion (Berry and Berry 2007; Jones and Baumgartner 2005; Shipan and Volden 2012), are still awaiting in-depth comparative inquiry.

Qualitative experimental designs, like qualitative comparative analysis (QCA), so far were only addressed in the *JCPA* by Rizova (2011), and a special issue edited by Brans (2017). The same trends appear to be the case in the EBSCO comparative articles. Fischer and Maggetti (2017), for instance, recently made the case for the usefulness of the controversial qualitative comparative analysis; large-N qualitative methods are used more rarely than case studies or statistical analysis.

The addition of a special section devoted exclusively to statistics in 2014 in the *JCPA* has helped to increase the prevalence of quantitative approaches. However, these studies more often rely on publicly available data, mostly from national governments or multinational organizations like the OECD and EU, rather than on data generated by surveys designed by the researchers. A few exceptions to this rule, in the *JCPA*, are Benito and Brusca (2004), Avrami and Rimmerman (2005), and Varma and Kapur (2013).

There remains a dearth of work that employs comparative interview or ethnographic methods. Although these methods are not completely absent from the EBSCO database journals (8%), on the pages of *JCPA* they show only 2%. They are exemplified by the recent work of Maybin (2015) and Escobar (2015) in a special issue on “Professional Knowledge and Policy Work” (*JCPA*, 17 (3)) which attempted to pursue inquiry into how policies are subjectively perceived vis-à-vis objective measures. More research based on interviews and ethnography would contribute to the robustness of the field.

In the face of globalization and international policy dependencies, many policy journals in recent years have published articles with an international comparative policy scope. The *JCPA* promised to include articles about five continents and has adhered to this commitment. Notwithstanding, the developing world is studied far less than the developed countries both in the comparative EBSCO database articles, and in the *JCPA*. The predominance of authors from the developed world likely reflects the nascent status of the field of comparative policy analysis and comparative policy analysis studies (Muhleisen and Mukherjee 2016). Increasing

awareness in comparative policy analytic work means that EBSCO journals promoting comparative studies and the *JCPA* should seek to encourage more studies from Central Asia, the Middle East, Africa, Central America, and the Caribbean. To expand *JCPA*'s geographic reach, the *JCPA* has published a special issue on Policy Analysis in Eastern Europe, and another special issue is forthcoming on comparative public policy experimental methodology based on a recent workshop at Tsinghua University, Beijing. It also initiated the Best Comparative Paper Award at The Network of Institutes and Schools of Public Administration in Central and Eastern Europe (NISPAcee) for papers presented at the NISPAcee annual conference and *JCPA* workshop, at the University of Iasi, Romania. EGPA's two policy groups have partnered to grant the award for the best comparative paper, as have APSA, APPAM, NASPAA, MPSA, IPISA, and the Chinese Public Management Academy. Papers awarded or presented are eligible for publication in the *JCPA*, following standard review.

Interestingly, we note a certain fit between geographical and methodological trends. The presence of open public data and independent journalistic and academic establishments is likely to be correlated with democracy and GDP. Indeed, the developed nations that tend to be studied most frequently, also tend to have relatively open governments that produce reliable data for quantitative analysis and keep publicly accessible records; they have free presses that report on policy developments with some degree of objectivity, and they have developed academic establishments to coordinate and oversee research. Consequently, historical, case-based, qualitative and quantitative methods can be used with relative ease. Ethnographic methods might be relatively more useful in countries where good government-generated data is not available, although they certainly are important even in developed countries.

As regards disciplines, the content analysis has indicated that both the EBSCO articles and the *JCPA* have a strong record of comparative policy publications in the areas of economic and financial policy, welfare and social subsidies, and health. It is fairly strong on international agreements and organizations, the environment, education, and immigration/migration. But there are many important questions of policy that fall beyond this scope. Christos Kassimeris's (2006) effort to advance a framework for comparing foreign policies is one of just a handful of articles on this topic. Dicle and Dicle (2010) study military spending, but there are few comparative studies on the outcomes of military intervention. There is no work on the effectiveness of economic and other sanctions as a tool of international diplomacy. Little comparative scholarly work has been published on human rights public policies although some fundamentals are included in studies in social welfare, healthcare, and environmental policies. Montefrio (2014) and Golder and Williams (2006) are part of a small number who do make contributions in this area. There have been few studies on crime and criminal sentencing, prison and incarceration, and racial bias and discrimination, despite a wealth of raw data related to policies in these fields.

Comparative policy studies have a central role in today's international policy making and social and economic development. The scholarly community has an important function in advancing the comparative policy dialogue to address systematic, evidence-based, responsible policy making and facilitate effective and efficient optimal policy solutions. Through their comparative studies, comparative policy researchers are those able to extend this field with the necessary robust scientific scholarly cornerstones by continuing to pursue comparative pioneering work. At this stage of its development as a new domain of social sciences, comparative policy studies necessitates attention being paid to comparative methodology and comparative

theory directly pertinent to public policy, as well as to less explored disciplines and geographic regions.²

NOTES

1. Google's description of the Google Million: "The 'Google Million'. All are in English with dates ranging from 1500 to 2008. No more than about 6000 books were chosen from any one year, which means that all of the scanned books from early years are present, and books from later years are randomly sampled. The random samplings reflect the subject distributions for the year (so there are more computer books in 2000 than 1980)." Retrieved from <https://books.google.com/ngrams/info> on October 8, 2016.
2. A version of this chapter was published in *JCPA*, Vol. 20 (1) (2018), a special issue devoted to the JCPA's 20th anniversary.

REFERENCES

- Avrami, S. and Rimmerman, A. (2005). Voting intentions of Israeli legislators regarding proposed disability and social welfare laws. *Journal of Comparative Policy Analysis: Research and Practice* 7 (3): 221–32.
- Béland, D. and Howlett, M. (2016). The role and impact of the multiple-streams approach in comparative policy analysis. *Journal of Comparative Policy Analysis: Research and Practice* 18 (3): 221–7.
- Benito, B. and Brusca, I. (2004). International classification of local government accounting systems. *Journal of Comparative Policy Analysis: Research and Practice* 6 (1): 57–80.
- Berry, F. S. and Berry, W. D. (2007). Innovation and diffusion models in policy research. In P. A. Sabatier (Ed.), *Theories of the Policy Process* (2nd edition). Boulder, CO: Westview Press, 169–200.
- Brans, M. (2017). Introduction. *Journal of Comparative Policy Analysis: Research and Practice*, 19 (4).
- Brans, M., Geva-May, I., and Howlett, M. (2017). Policy analysis in comparative perspective: introduction. In M. Brans, I. Geva-May, and M. Howlett (Eds.), *Routledge Handbook of Comparative Policy Analysis*. New York: Routledge, 1–24.
- Clavier, C. (2010). Bottom-up policy convergence: a sociology of the reception of policy transfer in public health policies in Europe. *Journal of Comparative Policy Analysis: Research and Practice* 12 (5): 451–66.
- Coleman, W. D. (2001). Agricultural policy reform and policy convergence: an actor-centered institutional approach. *Journal of Comparative Policy Analysis: Research and Practice* 3 (2): 219–41.
- deLeon, P. and Resnick-Terry, P. (1998). Comparative policy analysis: déjà vu all over again? *Journal of Comparative Policy Analysis: Research and Practice* 1 (1): 9–22.
- Dicle, B. and Dicle, M. F. (2010). Military spending and GDP growth: is there a general causal relationship? *Journal of Comparative Policy Analysis: Research and Practice* 12 (3): 311–45.
- Dierkes, M., Weiler, H. N., and Antal, A. B. (1987). *Comparative Policy Research: Learning from Experience*. Aldershot: Gower.
- Escobar, O. (2015). Scripting deliberative policy-making: dramaturgic policy analysis and engagement know-how. *Journal of Comparative Policy Analysis: Research and Practice* 17 (3): 269–85.
- Fischer, M. and Maggetti, M. (2017). Qualitative comparative analysis and the study of policy processes. *Journal of Comparative Policy Analysis: Research and Practice* 19 (4): 345–61.
- Gerring, J. (2006). *Case Study Research: Principles and Practices*. New York: Cambridge University Press.
- Geva-May, I. (2002a). Cultural theory: the neglected variable in the craft of policy analysis. *Journal of Comparative Policy Analysis: Research and Practice* 4 (3): 243–65.
- Geva-May, I. (2002b). Comparative studies in public administration and public policy. *Public Management Review* 4 (3): 275–90.

- Geva-May, I. (2004). Comparative policy analysis studies: relevance and methodological principles. Colloquium Presentation, Concordia University.
- Geva-May, I. (2008). Comparative public policy studies. Presentation, Shih Hsin University, Department of Public Policy and Management, Taipei, Taiwan (May).
- Geva-May, I. (2015). Comparative public policy methodologies. Congreso Latinoamericano y Caribeño Ciencias Sociales, Quito, Roundtable (August).
- Geva-May, I. and Lynn, L. E. Jr. (1998). Comparative policy analysis: introduction to a new journal. *Journal of Comparative Policy Analysis* 1 (1): 5–8.
- Geva-May, I. with Wildavsky, A. (1997). *An Operational Approach to Policy Analysis: The Craft*. Boston: Kluwer Academic Publishers.
- Golder, B. and Williams, G. (2006). Balancing national security and human rights: assessing the legal response of common law nations to the threat of terrorism. *Journal of Comparative Policy Analysis: Research and Practice* 8 (1): 43–62.
- Happaerts, S. and Van Den Brande, K. (2011). Sustainable development and transnational communication: assessing the international influence on subnational policies. *Journal of Comparative Policy Analysis: Research and Practice* 13 (5): 527–44.
- Heady, F. (2001). *Public Administration: A Comparative Perspective* (6th edition). New York: Dekker.
- Ingraham, P. (1996). *A Laggard's Tale: Civil Service and Administrative – Building a Government that Works*. Washington, DC: Brookings Institution Press.
- Jones, B. D. and Baumgartner, F. R. (2005). A model of choice for public policy. *Journal of Public Administration Research and Theory* 15 (3): 325–51.
- Jreisat, J. E. (2002). *Comparative Public Administration and Policy*. Boulder, CO: Westview Press.
- Kassimeris, C. (2006). A depiction of the foreign policy behaviour of dyadic case studies based on the “Triangle Model”. *Journal of Comparative Policy Analysis: Research and Practice* 8 (3): 271–81.
- Kuipers, S. (2009). Paths of the past or the road ahead? Path dependency and policy change in two continental European welfare states. *Journal of Comparative Policy Analysis* 11 (2): 163–80.
- Lasswell, H. D. (1968). The future of the comparative method. *Comparative Politics* 1 (1): 3–18.
- Lasswell, H. D. and Kaplan, A. (1950). *Power and Society: A Framework for Political Inquiry*. New Haven, CT: Yale University Press.
- Lindblom, C. E. (1958). Policy analysis. *American Economic Review* 48 (3): 298–312.
- Maybin, J. (2015). Policy analysis and policy know-how: a case study of civil servants in England's Department of Health. *Journal of Comparative Policy Analysis: Research and Practice* 17 (3): 286–304.
- Mill, J. S. (1843). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. London: John W. Parker.
- Montefrio, M. J. F. (2014). State versus indigenous peoples' rights: comparative analysis of stable system parameters, policy constraints and the process of delegitimation. *Journal of Comparative Policy Analysis: Research and Practice* 16 (4): 335–55.
- Muhleisen, J. and Mukherjee, I. (2016). Policy analysis: a rich array of country and comparative insights. *Asia Pacific Journal of Public Administration* 38 (3): 204–10.
- Neustadt, R. E. and May, E. R. (1986). *Thinking in Time: The Uses of History for Decision Makers*. New York: Simon & Schuster.
- Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action. Presidential Address, American Political Science Association, 1997. *American Political Science Review* 92 (1): 1–22.
- Radin, B. A. (2013a). *Beyond Machiavelli: Policy Analysis Reaches Midlife* (2nd edition). Washington, DC: Georgetown University Press.
- Radin, B. A. (2013b). Policy analysis reaches midlife. *Central European Journal of Public Policy* 7 (1): 8–27.
- Rizova, P. S. (2011). Finding testable causal mechanisms to address critical public management issues. *Journal of Comparative Policy Analysis: Research and Practice* 13 (1): 105–114.
- Rose, R. (1991). What is lesson-drawing? *Journal of Public Policy* 11 (1): 3–30.
- Rose, R. (1993). *Lesson-Drawing in Public Policy: A Guide to Learning across Time and Space*. Chatham, NJ: Chatham House.
- Rose, R. and Mackenzie, W. J. M. (1991). Comparing forms of comparative analysis. *Political Studies* 39 (3): 446–62.

- Shipan, C. R. and Volden, C. (2012). Policy diffusion: seven lessons for scholars and practitioners. *Public Administration Review* 72 (6): 788–96.
- Swanson, G. (1971). Frameworks for comparative research: structural anthropology and the theory of action. In Ivan Vallier (Ed.), *Comparative Methods in Sociology*. Berkeley, CA: University of California Press, 141–202.
- Varma, R. and Kapur, D. (2013). Comparative analysis of brain drain, brain circulation and brain retain: a case study of Indian Institutes of Technology. *Journal of Comparative Policy Analysis: Research and Practice* 17 (4): 315–30.
- Weimer, D. L. (1993). The current state of design craft: borrowing, tinkering, and problem solving. *Public Administration Review* 53 (2): 110–20.
- Weimer, D. L. and Vining, A. R. (1999). *Policy Analysis: Concepts and Practice*. Englewood Cliffs, NJ: Prentice Hall.
- Wildavsky, A. (1979). *Speaking Truth to Power: The Art and Craft of Policy Analysis*. Boston, MA: Little, Brown.

22. Evolutionary theory in comparative policy analysis

Adrian Kay

1. INTRODUCTION

The prospects for evolutionary theory in comparative public policy, as well as neighbouring fields in public administration, have attracted attention in recent decades (e.g. Bovaird 2008; Cairney 2012, 2013; Dowding 2000; Haynes 2008; John 1999, 2000; Kerr 2002; Klijn 2008; Lewis and Steinmo 2010; Lustick 2011a, 2011b; Steinmo 2010; Tenbensel 2013, 2015, 2018). This attention can claim a long intellectual history in the social sciences; the argument that we can think of societies in terms of evolution goes back to the development of social sciences in nineteenth-century German-speaking universities and a set of claims that, even without any strong analogy between the natural and social worlds, it can be useful to structure arguments about social phenomena as evolutionary analysis (Hodgson 2002).

This chapter considers several methodological issues in the use of evolutionary theory to warrant descriptions, explanations, or interpretations in the study of public policy. It offers reflections on the prospects and pitfalls of evolutionary theorizing in the comparative study of public policy. Although there is no consensus in the literature about practical prospects for evolutionary approaches, it has revealed how certain key theories in policy studies are explicitly evolutionary in their inspiration, and that there is a larger set of theories that are at least consistent with evolutionary approaches to public policy.

This chapter argues that this recognition can support something more than a simple relabelling exercise and that there are at least two methodological advantages to recognizing evolutionary elements in prominent public policy theories. The first is that evolutionary approaches facilitate theoretical triangulation. The chapter will show how evolutionary approaches can group institutional theories in public policy by their logics of inquiry rather than disciplinary origins, or their modelling preferences. Evolutionary thinking does not produce another institutionalism or relabelling of existing institutionalisms but rather a framework for sorting existing theory, with the advantage of providing the means for comparative assessment of the usefulness of their different models in policy studies. For example, evolutionary analysis opens lines of inquiry for agency in institutionalism in which it is habits and norms that attach to groups rather than fully formed policy ideas as, for example, in the notion of coalition magnets (Béland and Cox 2016). This helps *inter alia* with questions of the institutional formation and institutional origins in self-organization rather than design processes.

The second benefit is in methodological support for certain methods already established in the study of public policy. The use of evolutionary theory for description, interpretation, and explanation encourages the sort of comparative case study research designs as well as in-depth historical narratives already widely used in comparative policy studies. Many public policy events, episodes, and processes are not readily subsumed under general analytical categories for comparison, they are instead unique. Evolutionary theory acknowledges this but suggests

that in cases which are non-recurrent, contingent, and conjunctural, the causal mechanisms we may wish to draw inferences about do operate across time and space. In other terms, though a policy episode may be unique, explained by a highly contingent conjunctural combination of causal mechanisms, those mechanisms themselves are comparable.

This chapter seeks to make contributions to complement and extend this line of scholarly inquiry in the methodology of comparative public policy. The chapter is organized into three steps. First, it investigates the distinction between evolution as a process and evolution as a theory in greater depth than the current policy studies literature. In doing so, we argue that the claim that evolution can be used as a metaphor is more important methodologically than is currently acknowledged. In particular, evolutionary thinking as a metaphor is related to analytical strategies and prior to the analogy, concept formation and comparison that underpin subsequent application of evolutionary theory in comparative public policy.

Second, the chapter argues that evolutionary theory can be stated more usefully at a higher level of abstraction than existing comparative public policy theories. In methodological terms, it can serve as part of a framework of open systems thinking about policy that can help comparisons of the advantages and the disadvantages of extant policy theories and models. This claim is illustrated with respect to the use of Pragmatist approaches to institutional foundations in comparative policy analysis.

In the third part of the chapter, the regularly observed methodological problems of an evolutionary approach in the social sciences are set out; the links to functionalism, the role of agency, and the problem of circularity. Partial answers to these disadvantages are provided in terms of evolutionary approaches in the methodology of comparative public policy.

2. EVOLUTIONARY APPROACHES

Although the terms ‘evolution’ and ‘evolutionary’ are widely used in the social sciences, for the most part they are employed as convenient shorthands for slow and gradual change over time. To introduce evolutionary thinking into policy analysis, evolution can usefully be distinguished as a descriptive label for a process of change from biological evolution as a theory of change in the natural world (Cairney 2013; John 1999; Kerr 2002). Instead, evolutionary perspectives are something useful for the analysis of all open, complex systems, of which natural systems are an important but not unique example (Campbell 1965). This allows the analogy to be made between the problems of evolutionary theory in explaining natural processes of change and the problems of evolutionary theory explaining processes of social change in terms of issues such as time, history, and differential rates of change (Bardach 2006; Kay 2006).

2.1 Evolution as Process

Gradualism has often been understood as a defining element of evolution as a process (Campbell 1997). At the heart of the Darwinian view of evolution is gradualism; mutations to the genome can cause minor variations in the organism’s properties. These variations add up over generations and generations to create the complex organisms observed in nature.

In comparative public policy, there is a connection between gradualism in a process of change and the reductionism observed in several theories of policy change. If policy, understood as a composite whole, changes only gradually it is easy to view it as the sum of

relatively independent parts, with the corollary that these parts may be analysed independently without the need to consider significant interaction effects on the whole. For example, we can study policy instrument choices, combinations, and adjustments without reference to the higher order policy paradigm within which they operate (Béland and Howlett 2016; Kay and Daugbjerg 2015).

This is the *ceteris paribus* method employed in simple neo-classical economics. However, if we assume that every component of policy is strongly connected to all other components, a minor change in one component influences all other components. The density of interconnections among elements is a measure of the complexity of a system (Klijn 2008). When this is high, and the interactions are at some speed, then we may argue that complexity in policy systems makes the assumption of gradualism difficult to sustain as an analytical strategy.

Reductionism is an important and consistent strand in policy studies, where policy is conceived as either a macro- or meso-scale phenomenon that can change incrementally and be reduced analytically to the behaviours of independent micro-level units of analysis, such as policy makers. Thus, gradual changes in policy are understood in terms of the bounded rationality of policy makers or institutionally constrained agents (Cairney 2013).

However, the assumption of gradualism is problematic for understanding evolution as a process where there is observed stability as well as different rates of evolution. A controversial and long-standing question in evolutionary biology is whether natural history contains emergent evolutionary changes or novelties that are not merely the accumulation of small, gradual, and adaptive steps. In a celebrated work, Eldredge and Gould (1972) looked at fossil records and found that morphological characters stay the same for very long periods, occasionally punctuated with drastic change within a short period of time.

The assumption of gradualism is similarly difficult in the policy-making world. The idea of *punctuated change* has had a strong grip on thinking about evolution as a process in the political world (Cairney 2013). For example, workhorse models of the dynamics of public policy by Kingdon (1995) and Baumgartner and Jones (2002, 2009) have used the notion of punctuated change; there are moments, windows, or critical junctures where change is observed. Although these do not represent a full evolutionary theory or a detailed and contextualized application of evolutionary concepts in the social sciences, they do acknowledge their borrowings from evolutionary biology.

The more general point, for this chapter, does not concern the difference between gradual or punctuated change, rather whether interactions between different units of comparative public policy can produce evolutionary effects in terms of *self-organization* rather than the gradual grinding of selection mechanisms. Within any system that has repeated interactions among its constituent elements with feedback (both positive and negative) and agents capable of innovation, there is the possibility that the internal order of a policy system can increase. Within the developing policy feedback research agenda (Moynihan and Soss 2014), this notion of policy as an emergent property of self-organization rather than top-down design still lacks sustained attention. In this strand of evolutionary thinking, policy is arrived at rather than made, is independent of external selection pressures, and does not rely on an assumption of some form of agency operating at the system level.

If we admit such a possibility then the perspective of evolution as a process moves away from approximating policy as the sum of independent actions by policy makers. Kauffman (1995) uses the idea of complexity to doubt gradualism in evolutionary processes in the natural world by establishing that in some complex systems minor mutations cause significant

changes in the system due to interaction effects. The influence of Kauffman's work is manifest in a study of international politics by Jervis (1999) as well as work in evolutionary economics (Dopfer 2012; Potts 2000, 2014). All are concerned with self-organization: how patterns of collaboration and cooperation can emerge quickly, potentially spontaneously, out of the interactions of agents within certain complex systems (Lustick 2011a).

As a possible future direction for policy studies in approaching self-organization, evolutionary economics is distinguished by its analytical ability to handle simultaneously the effects of individual choices and the selection pressures from the system (Weise 1996). There would be no 'either/or' rivalry between rational choice on the part of the individuals and institutional selection pressures in the policy system; instead analysis concentrates on their concatenation in generating a policy path. For Weise (1996), individuals may coordinate their behaviour through incentives and sanctions attached to formal institutions but also, less formally, through habits and norms forms of cooperation. This speaks to the nascent literature on bricolage in policy studies (Baker and Nelson 2005; Carstensen 2011; Freeman 2007; Wilder and Howlett 2014). Bricolage is an evolutionary process in which policy is negotiated as opposed to designed. International trade treaties or omnibus reform bills are examples of hybrid and synthetic policy designs, and the claim is that these are typical of policy more broadly. There is no coherent policy paradigm to inform policy designs, or ideas to act as coalition magnets (Béland and Cox 2016); instead individual agents creatively stitch together bits and pieces of ideas and other aspects of institutional legacies in order to negotiate misalignments between different policy instrument packages and policy goals.

The self-organization insight brings a feedback loop into this bricolage literature. The interaction and relations among agents can make the policy 'whole' more than the sum of its parts. The interactions and relationships in a policy system form institutions, generate variety, and produce complexity in the interaction of ideas, interests, and material circumstances. This makes composition important: these relations cannot simply be summed; rather they are dynamic, complex, and typically have emergent, self-organizing properties. This is missed by the three orders or levels of change from classifications of policy change following Hall (1993), and subsequent evolutionary-inspired work (e.g. Cashore and Howlett 2007).

This challenge to reductionism is different from the emphasis over the last decade in the new institutionalism on strategic agency and the need for 'firm' micro-foundations for theories of institutional and policy change. This is driven by the valid concern that institutional analysis should not 'overdetermine' behaviour or actions. The potential for creative bricoleurs and entrepreneurial agency means institutions cannot be explained uniquely and cannot simply be assumed to persist indefinitely. This is not the same as supporting methodological individualism in policy studies. By always focusing on the individual agent and how their decision making is affected by institutional structure, evolutionary effects in terms of the emergent properties of composite wholes are liable to be missed. Institutions are collective: it is groups of agents following a rule that constitute an institution. Significant interaction effects include tipping points, network effects, combinational effects, bandwagon effects, reinforcement, emergence, and learning and imitation. These are all concepts from the evolution of complex systems that can be used to structure accounts of policy processes without requiring any assumption of individual agency.

Further, by acknowledging that there are properties of the whole that are not reducible to its constituent elements, the analytical possibility of macro- and meso-level evolutionary effects is raised. This is where the causal mechanism runs from macro-level to macro-level rather

than macro- to micro-. For comparative public policy purposes, this means causal mechanisms operating at a level higher than individual agency. Where, for example, policy change can cause policy change without individual agency being a key explanatory feature. For example, many studies of the regulatory state or meta-governance are observations of meso-level evolutionary effects (e.g. Moran 2002; Dubash and Morgan 2013).

2.2 Evolution as Theory

Any evolutionary theory supports a form of consequence explanation in the sense that phenomena are explained through their actual consequences. The central task for an evolutionary theory supporting such explanation-by-consequences is to provide a mechanism by which the consequences uphold or maintain the action or structure one wants to explain (Cilliers 1998). In the absence of some kind of feedback from effect to cause, explanation by consequence remains unclear. A functional explanation (FE) is a special class of consequence explanation where the consequences of an institution or routine of behaviour are favourable or functional, for some agent or group who maintain that institution or behaviour. Explaining a cause, a functional institution X, in terms of its consequences, the functions Y it carries out, reverses the temporal sequence of causes preceding consequences.

In terms of temporality, this is a notably different approach to the co-variational research designs that are used conventionally in comparative public policy. Not least in reversing the temporal sequence in the ‘causes-of-effects’ approach (Goertz and Mahoney 2012), where an empirically predicted co-variation between X and Y across different cases alongside theorized and observed causal mechanisms in those cases underpins claims of causation (Peters, Fontaine, and Mendez 2018).

The famous variation-and-selection model of Darwin, labelled subsequently as the ‘survival of the fittest’ by Herbert Spencer, says that functional structures will lead to greater survival and replication of the genes that produce them. However, within the social sciences, the traditional riposte has been the ‘missing mechanism’ argument used more generally to deny all ‘biological analogies’ in the social world. That is, there is simply no equivalent of natural selection operating in the social sphere, still less a rough equivalent in comparative study of public policy processes.

Usually the answer is to emphasize historical contingency, accident or random variation in terms of why an X occurs at some time; then focus on feedback loops from Y to X that explain why X continues to perform the function (Elster 1983). The separation of analysis between the creation of institutions and what sustains them is a characteristic of historical institutionalist analysis (e.g. Mahoney and Thelen 2010; van der Heijden 2010).

In terms of thinking about evolutionary thinking in policy analysis, natural selection as a feedback loop successfully supports a functional explanation. Elster (1983) discusses why functionalism is a good strategy in biology by presenting the traditional variation-and-selection model, which says that functional structures will lead to greater survival and replication of the genes that produce them. Cohen (1978) notes the following Marxist argument as a FE: the bourgeois media report industrial conflict in a style which favours the capitalist class because that style of reporting has that tendency. But Elster (1980) argued that even though a FE remained logically possible, no mechanisms to support one actually exist. This is the missing mechanism argument often used to deny ‘biological analogies’ in the social world. There

is simply no equivalent of natural selection operating in the social sphere. FEs in the social sciences can only work if supported by evidence of institutional selection.

Campbell (1965, 1974) argued that Darwinism contained a general theory of the evolution of all complex systems of which organic evolution was only one. This argument is prevalent in contemporary evolutionary economics. For example, Hodgson (2002: 270) argues that ‘there is a core set of general Darwinian principles that, along with auxiliary explanations specific to each specific domain, may apply to a wide range of phenomena’ (also see Dopfer 2012; Dopfer and Potts 2007; Potts 2000). Nelson and Winter (1982) applied the principles of variation, inheritance, and selection to routines in firms. Additionally, a tradition of evolutionary epistemology stretches back to nineteenth-century Pragmatism. More recently, this has included work on experimentalism governance which traces the development of policy useful knowledge in terms of experimentation, trial and error learning, and selection (Sabel and Zeitlin 2012).

In asserting a general set of principles for the analysis of all complex systems, evolutionary theory amounts to a claim of the existence of processes of variation, retention, and selection irrespective of the particular mechanisms that might be plugged in for specific theoretical questions such as policy change.

In such terms, and against simply assessing evolutionary theory as consistent with other policy theories (Cairney 2013), the argument is that evolutionary theory is not just another theory stated at a similar level of abstraction as workhorse policy theories but should be judged at a higher level of abstraction, able to be employed methodologically to compare and contrast the utility of existing theories in terms of open, complex systems. This still leaves the sceptic’s question: what is the relationship of these general evolutionary principles to acceptable theoretical presuppositions about the policy-making world or empirically based understandings of policy making?

The challenge is not to uncover a mechanism of selection that is the equivalent of genetic selection in the natural world because the argument no longer proceeds by analogy from the natural to social worlds. Instead, evolution requires a search for mechanisms of variation, retention, and selection that are appropriate and contextualized in the policy-making sphere.

2.3 Evolution as Metaphor

The previous section argued that evolutionary theory may be employed for both gradual and abrupt policy change and contribute to understanding policy development in terms of self-organization, emergence, and selection pressures. In doing so, evolution may act as a metaphor in policy studies. The claim that frameworks or theories act as metaphors often implies weakness and a lack of substance. However, as recent philosophers of social science have noted, metaphors can have a deeply constitutive, if subterranean presence, in the formation of concepts and analysis of change (Klamer and Leonard 1994; Lakoff and Johnson 2003; Lewis 1996). Indeed, metaphors are prior to any analogy; they establish broad and general mappings across conceptual domains and help structure understanding and assist in the perception of connections between different things. In turn, this supports the formation of concepts to organize things into different categories (Lakoff and Johnson 2003).

In terms of extant studies of policy dynamics, evolutionary metaphors are a useful way of organizing thinking about policy learning by distinguishing adaptation (learning that affects calculations about how to realize interests most effectively, sometimes called Bayesian ration-

ality) from more complex learning where interests, identities, and institutions are learned or constructed in the interaction of agents in the policy system. In this latter sense of learning, the problem situation itself is constructed in the interactions between agents. Policy problems, policy solutions, and criteria of ‘success’ (or the intentional selection mechanisms that operate to ‘weed’ out policy failure) are all constructed (Marsh and McConnell 2010). It is through this construction that policy paradigms emerge and are institutionalized (Carstensen 2017; Wilder and Howlett 2014). As noted, Kauffman (1995) argues that analysis of the internal dynamics of self-organization should complement consideration of selection mechanisms in evolutionary theory. In this account, evolution is not just ad hoc fiddling, but that bricolage drives self-organization, the possible emergence of order, upon which selection mechanisms operate to support or retard development and subsequent prospects for its institutionalization (Jabko and Sheingate 2018).

In the new institutionalism research programme, influential in comparative public policy, there is the potential for institutional change in terms of conversion, recombination, layering, drift, and activating redundancy (Béland 2010; Crouch and Farrell 2004, Mahoney and Thelen 2010; Streeck and Thelen 2005; van der Heijden 2010). These are all examples of evolutionary change as self-organization that may or may not become institutionalized over a period through environmental selection pressures.

The evolutionary metaphor also helps suggest the idea of rationality as contextualized and emerging from the interactions of agents in the policy system (Hoppe 2002). This is an important contribution to policy studies; it is not a rejection of the importance of rationality in policy making per se but rather a rejection of a unique and universal view of rationality and an emphasis on a plurality of rationalities that compete within the policy system (Spicer 2014; Wagenaar 1999). As much of the writing in the critical policy studies research seam argues, some rationalities in policy carry assumptions that are strict and universalizing, others more participative and contextual.

At first glance, the evidence-based approach to policy advocated by many governments around the world seems to assume the neutral and scientific treatment of evidence to inform policy intervention: an impartial and scientific rationality (Head 2016). Yet the evidence base also informs policy in terms of ‘what works’, a Pragmatist conception that comes as much from a particular context as it does from the application of scientific principles. The significance of the particular context has been heightened in the debates in other policy contexts over questions of recognizing identity and diversity.

In evolutionary biology, the process of natural selection operates at a population level. This has been criticized as ‘beanbag genetics’, involving circular arguments about population statistics and criticized for relying on comparative static models that are unable to account for actual processes of change (Dronamraju 2010). The evolutionary approach in policy studies shares this characteristic with evolutionary biology (Shpaizman 2017). Much of the best practice in case selection in comparative public policy relies on identifying the relevant population of cases. However, when the population is unknown, a situation common to much of comparative public policy where the population is less visible, these case selection guides are much less useful. As Shpaizman (2017) argues, research tends to either side-step this problem or, when recognizing it, default to a single case study research design.

In adopting a detailed, contextual scale of description of comparative policy development, public policy relies on narrative for explanation where mechanisms are expressed as tendencies, dispositions, and environmental limits combined with an emphasis on conjunctural

contingency, memory, and history. The evolutionary approach, while not at all ruling out intentional explanation and strategic agency, certainly decentres the agent in the policy process. The metaphor puts self-organization and selection mechanisms alongside intentional action in structuring narratives. Further, the evolutionary metaphor helps to introduce policy learning in a constructivist sense, which holds potential for understanding the emergence and institutionalization of policy paradigms (Carstensen 2011, 2017).

3. EVOLUTIONARY FRAMEWORKS FOR INSTITUTIONALIST THEORY IN COMPARATIVE POLICY ANALYSIS

The metaphor of evolution is prevalent, although not always acknowledged, in historical institutional analysis of continuity and change (Lustick 2011a; Steinmo 2010; Streeck and Thelen 2005; Thelen 2004). It is more explicitly referenced in classical Pragmatist conceptions of institutional development (Nungesser 2017). Although there is variation within classical Pragmatism as well as in its contemporary forms, an evolutionary perspective is common to all and contributes to institutional analyses by specifying *habits* as a central aspect in the functioning of institutions (Jabko and Sheingate 2018).

A Pragmatist viewpoint argues that ‘habits in action’ are the ‘flywheel’ of institutions and key to understanding processes of institutional formation and ongoing institutional evolution (Gronow 2011). Inspired by evolutionary theory, classical Pragmatists viewed habits not simply as individual capacities, but rather social-cultural phenomena continually formed in evolving transactions between organisms and their environment (Dewey 1948 [1922]; James 1950 [1890]). Habits precede individual beliefs or ideas about action; they are the units of inheritance for cultural contexts. Habitual action, understood as a predisposition to act, and its relationship to impulse and deliberation is at the core of the Pragmatist account of evolution in the social world. Instead of intentional design, where actors form intentions to act in the mind and then decide how to act and then act, in Pragmatism motives and preferences enter an ongoing process of action, or habits, as humans interact with their social environment (Whitford 2002). While actors may recombine their habits and learn to acclimatize or adapt to new situations, Pragmatists conceive of habits as always maintaining some element of continuity enabling action to proceed in unfamiliar environments. Habits are then the informal stock of socio-cultural predispositions and memories available for future problem solving in evolutionary economics (Nelson and Winter 1982) as well as underpinning the recent work on bricolage in comparative public policy.

The Pragmatist conception of habit encourages intelligent assessment of present opportunities for social change with reference to the past and potential future (Dewey 1946 [1927]). Routines, dispositions, and how rules operate in practice offer a window into the existing stock of habits in any given policy-making context. In situated action, embodied Pragmatist habits present actors with an opening for experimental responses: variety, deliberation and selection, or rather actively learning in interaction with the environment to respond (Sabel and Zeitlin 2012). Individual efforts to align new situations with actions and expectations present opportunities for ‘intelligence’ and reflection on the habits in use (Dewey 1948 [1922]). These moments of doubt or the demands for action in context require an adjustment of paradigms or new policy frames. Efforts by actors to reconstruct the situation to enable action must then test out the success of some new formulation in interaction with the social world. This includes

discursively constructing new possibilities for action but also seeking to establish institutional change in meaningful intersubjective experience. While Dewey held out the possibility that intelligent habits could transform institutional orders, he also recognized that power and coercion, or rather congealed habits, can prevent change (Ansell 2011; Dewey 1946 [1927]; Jabko and Sheingate 2018).

From an institutionalist perspective, Pragmatist concepts of habit are useful in providing an account of how institutional evolution is constituted by, but not reducible to, the creativity of actors involved in an ongoing process of habitual action (Jabko and Sheingate 2018; Kerr 2002). For Pragmatists, the means–ends relationship of policy goal and administrative action is never stable but evolving as actors learn to update the ‘means’ adopted to maintain the status quo ‘ends’ in place, or alternatively constructing new ‘ends in view’ – ends adjusted by means based on a stream of experiences – to which revised habits are put. This is a wider view of instrumentalism than usually acknowledged in comparative public policy. Recognizing a key role for creative individuals’ coping responses in contributing to ongoing institutional order and change, new recombinations of habit constitute micro-dynamics for ongoing reproduction or transformation of existing institutional arrangements. Recombinations of habit are always contestable and require acceptance and reproduction by a broader population of actors within the institutional environment. At a policy instruments level, efforts to recombine habits, such as by actors within public organizations, can be understood in an evolutionary sequence that can be conceived as having different starting points, contextual qualities, and with different potential outcomes. Such outcomes include interactive effects bringing about self-organization in the environment for ongoing change, dynamics that experience negative feedback from the environment, or putting in place dynamics that protect the existing institutional status quo from disruption (see Crouch 2005; Jabko and Sheingate 2018).

In keeping with the influence of Darwinian evolutionary theory on classical Pragmatist accounts of habit, the evolutionary metaphor can be used to describe and analytically distinguish sequences in the formation of institutions and individual efforts at habitual recombination and learning (Campbell 1965; Nungesser 2017). Variation, within a social context, relates to the conglomeration of habits and dispositions of actors incorporated into the design and interactive order of a given policy. Diversity of actors, and related (means–ends) epistemic resources and materials within a policy process supply the habits that inform actions and decisions reached. Variety of habits in the environment of the broader institutional configuration generally exceeds habits incorporated into policy action. The notion of the variation of habits is a useful starting point to describe the inheritance of newly formed institutions, providing a window into the available stock of habits to be deployed for future action, but also a connection to past iterations in an institutional domain.

If habits can help gain leverage on institutional formation, then selection pertains to the criteria and rationale for adopting a given standard or maintaining an institution. In contrast to biological evolution, cultural or social evolution has no natural selection as such, although institutional theories thrive on notions of survival and reproduction based on some form of social ‘fitness’ (Kay 2006). These include dysfunctional or interest-based selections that are the consequence of power and coercion rather than effectively resolving collective problems (Kay 2006; Jabko and Sheingate 2018). In applying the concept of selection to social-cultural contexts, attention is drawn towards the range of contextual factors being brought to bear on winnowing out or excluding certain variation in the environment from the internal institutional order (Kerr 2002). Selection pressures are then the forces shaping or constraining

policy choices. There are different kinds of selection pressures constraining the boundaries of administrative action including creating selection bias, or alternatively more explicit and active selection by actors or groups to assert strategic responses (Kerr 2002: 334). The burden is upon the researcher to identify these selection pressures and the mechanism through which force is asserted. The argument here is that the workings of selection pressures are observed at a micro-level in the recombination of actor habits.

Finally, retention or rather learning in a social context is captured by concepts of adaptation and acclimatization in the evolutionary metaphor (Baines and Kay 2019; Kerr 2002). Like scholarship on institutional entrepreneurs, contemporary evolutionary scholarship recognizes the role of key actors in ordering and filtering selection pressures. Learning then includes recombining habits with new variety, but also in that process potentially changing one's identity to enable adaptation of habits and relatedly adaptation to selection pressures in the environment. This leaves open space for unintended consequences from the interplay of variation and selection; including queries about the intentionality of actors. Far from some pre-ordained institutionally conditioned direction, Pragmatists suggest public policy reflects the ongoing conglomeration of existing habits. These existing habits shape the functioning of institutions, including public policy instruments within institutional configurations, in a manner that evolutionary theory helps apprehend.

4. EVOLUTIONARY THEORY IN COMPARATIVE PUBLIC POLICY: METHODOLOGICAL ISSUES

4.1 Functionalism

Evolutionary theories in the social sciences have often been elided with criticisms of functionalism (Kincaid 1996). To circumvent this criticism requires making the assumption that general evolutionary principles operate, but not in a way that 'selects' behaviours or institutions that produce favourable consequences for some group or agent as that would amount to functionalism. This is the dilemma of wanting evolutionary theory to have some analytical value and avoid functionalism. Kerr (2002) asserts that institutions or behaviours or ideas that have consequences which are beyond the 'strict limits or "selective" pressures' (Kerr 2002: 351) set by the environment will not tend to occur. As a result, 'this forces individual actors or groups of agents to negotiate, and "adapt" to, the context in which they are situated'.

The basic argument is that because dysfunctional elements tend to disappear, persisting forms can be assumed to be adapted in the sense of not being dysfunctional. It amounts essentially to the assertion that non-dysfunctional institutions maintain themselves over time, because they do not transgress environmental limits or constraints. Some might see this position as vulnerable because there may be many non-dysfunctional alternatives to a given dysfunctional institution. Without the ability to say which of them will emerge, and at what time, rather than simply that one of them ultimately will, the predictive capacity of evolutionary theory is limited. This brings us back to the earlier point that in terms of explanatory power, evolutionary theory is anaemic without functionalism.

However, there is good reason for this lack of explanatory power. The ambition for dynamic theories of public policy is for theories, methods, and concepts to produce intelligible narratives that involve highly contingent combinations of factors, and not universal theory of all

policy change. In terms of evolutionary theory, there may be different selection pressures operating at different levels and over different time horizons. There is no single uniform and universal mechanism that uniquely selects and that can be accounted for by a covering law theory. Instead, there is a multitude of overlapping and potentially conflicting environmental pressures over the short, intermediate, and long term.

The notion of selection as environmental pressure that produces a disposition or a tendency is closer to contemporary understanding of selection in the field of complex systems and evolutionary economics (Dopfer 2012; Hodgson 2014; Potts 2000). In these terms, selection mechanisms are not universal fields that operate consistently over time but rather are often local, relative, and operate over a specific period. Thus within the environment of the policy system there are dispositions, tendencies, and constraints that limit what policy or policy proposal or idea or advocacy coalition (and so on) may be successful.

4.2 Agency, Intentionality, and Evolutionary Theory

In a critical response to John's (1999) arguments in favour of introducing evolutionary theories into political science, Dowding (2000: 75) makes a distinction between evolutionary explanations and intentional explanation, in that 'specifying any non-intentional selection mechanism by which policies are generated may be called "evolutionary explanation"'. This distinction immediately limits evolutionary thinking in policy studies. Intentional actions are set out as not part of evolutionary theory, which instead is reduced to conjunctural contingencies, unintended consequences, and environmental selection pressures in terms of mechanisms to make sense of events and processes in policy making.

One way to accommodate selection and intentionality is through the concept of artificial selection, although on the Dowding distinction this is not evolutionary. The essential characteristic of artificial versus natural selection is that humans manipulate the criteria or environment of selection (i.e. it is intentional selection). However, the introduction of intentionality in this way raises the question of rationality, the processes of cognitive and cultural evolution that explain why agents come to act in the way that they do.

An example of intentional selection is evidence-based policy making. 'What matters is what works' has been the rationale for vastly expanded budgets and research staff in government departments and agencies. Evidence-based policy making in its use of pilots, monitoring, performance measurement and full evaluations and its commitment to act on the feedback of 'what works' is (at least in an ideal form) a type of evolutionary epistemology. The search for what works explicitly assumes that policy should be developed in an evolutionary way accepting the insight from Simon (1978) and Braybrooke and Lindblom (1963) that there are strong limits on notions of a universal rationality in policy making. Indeed, myopia and the importance of random variation or designed experimentation are at the essence of evolutionary theory. This is what distinguishes it from other models of the policy process.

Intentional selection mechanisms have been introduced into policy theory; particularly in the voluminous body of work that pushes analysis beyond the stages ideal of policy making. In Kingdon's (1995) influential multiple streams approach, different policy solutions enter at the initiation stage of policy, often taken from the 'garbage can' or 'policy soup' and are fitted by policy entrepreneurs to problems that 'float by'. There are selection pressures that filter out or fail to select certain policies; from backbenchers in the legislature, from different departments, from the electorate, the media or policy networks. This version of a 'policy window' is a policy

environment with a confluence of different policy, political, economic, and social trends occurring over different temporal scales. At those times certain ideas and policies ‘may have their time’ (Kingdon 1995: chapter 6). Kingdon sets out three main categories that enhance the chances of an idea’s survival: technical feasibility; value acceptability; and anticipation of future constraints. The first is whether the idea is fully worked out and crucially whether there is a practical and achievable implementation plan. The second is concerned with the compatibility with the values of the policy community. This is the equivalent of the notion of a policy paradigm. Third is the anticipation of future constraints: the idea must be seen to have the potential to have acceptable budgetary costs and enjoy reasonable prospects of approval from politicians and the public (mass, activist, media). These are all intentional selection mechanisms.

However, for Kingdon these selection criteria only draw up a shortlist. Which particular idea emerges depends on a series of contingencies in which tipping points and bandwagons can play a part. Beyond a certain level of popularity, opposition to an idea ceases or alternatives fade. Everybody joins the bandwagon. This is the internal dynamics emphasized by the self-organization strand of evolutionary theory, which was discussed earlier in this chapter.

4.3 **Circularity**

The technical definition of fitness is the relative rate of change in the number of instances of a certain character within a population (Gould 2002). It has no significance itself, but only makes sense in the distribution of fitness for a population of organisms. Why do characters grow at different rates? Natural selection answers this question but raises the further issue: what are the causal mechanisms responsible for the variation in growth rates of different characters? Evolutionary biology as based on the ‘survival of the fittest’ has often been challenged on this question (Gould 2002). Fitness is defined in terms of those units that survive and reproduce successfully. In alternative terms, survival defines fitness. Therefore, how should we explain survival? According to natural selection, those that are the fittest and most capable of reproducing successfully survive. So fitness explains survival. This invites the criticism that evolutionary theory is vacuous or circular.

The practical consequence of this circularity is that evolutionary theory may be labelled descriptive in the sense of not allowing any counterfactuals or instances incapable of being falsified. In other terms, evolutionary theory is incapable of prediction either prospectively or retrospectively even in the statistical sense (Foster and Pyka 2014). This is important if you judge theory by its explanatory and predictive powers. It is less of a problem if you view the role of theory as contributing to historical explanation of particular policy cases. With policy dynamics there never are counterfactuals other than those imagined by scholars and the scale of description is detailed, contextual, and fine-grained.

5. **CONCLUSION**

The comparative turn in policy studies, even within its heterogeneity, disunity, and absence of overarching frameworks, has helped move scholarly inquiry slowly beyond a stability versus ‘big’ change dualism in its analysis of change. In doing so, a modest but noteworthy engagement with evolutionary thinking has emerged. Such engagement has helped support

conceptual work on both chance and selection mechanisms in the comparative analysis of policy change, but also reveals the methodological importance of the issue of functionalism in any evolutionary inspired study of public policy. The challenge is to develop full accounts of the political and policy consequences of dysfunctions between policy structures, their effects, and selection pressures operating in the broader policy environment.

More generally in policy studies, evolutionary approaches support understandings of policy as a complex whole, characterized at any particular moment by goals, objectives, and the calibrations of existing policy tools. In turn, policy reform is about changing certain elements of established or inherited policy whole. In any such change, there are a multitude of decisions with various effects but also the result of various influences. Some of these can be observed easily and others not. In policy studies, there is a tendency to assume implicitly there is a decision maker making these changes. However, drawing on the metaphor of an evolutionary process, this chapter has argued that this can lead to an underappreciation of the dynamics of policy systems and subsystems that are not the direct result of agency but rather the emergent property of policy as a whole. Of course, policy is constituted by actions, indeed agency, at different levels and parts of a policy system according to certain institutions (formal and informal) but at the system-level there is sometimes no actual policy change decision taken or action intended.

From an evolutionary perspective, discussions of successful policy reform strategies need to be supplemented by the acknowledgement that the target is moving; the policy system is changing – perhaps even reforming – through incremental accumulation of uncoordinated actions and not necessarily gradual consequences. This is why reform efforts in policy do not just confront inertia or stasis, as usually stated in certain strands of the new institutionalism literature, but also the problem of dynamic order.

The dynamic nature of changes in the policy environment and associated variations in the operation of policy selection mechanisms as well as the random errors in policy reproduction are not explicitly modelled in these newly freighted concepts to capture gradual but transformative change. The purpose of the case study presented is to establish an argument that they should be; because they redefine the nature of barriers to agent-centred, intentional reform strategies.

REFERENCES

- Ansell, C. (2011). *Pragmatist Democracy: Evolutionary Theory as Public Philosophy*. Oxford: Oxford University Press.
- Baines, D. and Kay, A. (2019). Evolutionary approaches to the concept of drift in policy studies. *Critical Policy Studies* 13 (2): 174–89.
- Baker, T. and Nelson, R. (2005). Creating something from nothing: resource construction through entrepreneurial bricolage. *Administrative Science Quarterly* 50 (3): 329–66.
- Bardach, E. (2006). Dynamics of public policy. In M. Moran, M. Rein, and R. Goodin (Eds.), *The Oxford Handbook of Public Policy*. Oxford: Oxford University Press, 336–66.
- Baumgartner, F. and Jones, B. (Eds.) (2002). *Policy Dynamics*. Chicago: University of Chicago Press.
- Baumgartner, F. and Jones, B. (2009). *Agendas and Instability in American Politics* (2nd edition). Chicago: University of Chicago Press.
- Béland, D. (2010). Reconsidering policy feedback. *Administration and Society* 42 (5): 568–90.
- Béland, D. and Cox, R. H. (2016). Ideas as coalition magnets: coalition building, policy entrepreneurs, and power relations. *Journal of European Public Policy* 23 (3): 428–45.

- Béland, D. and Howlett, M. (2016). How solutions chase problems: instrument constituencies in the policy process. *Governance* 29 (3): 393–409.
- Bovaird, T. (2008). Emergent strategic management and planning mechanisms in complex adaptive systems. *Public Management Review* 10 (3): 319–40.
- Braybrooke, D. and Lindblom, C. (1963). *A Strategy of Decision*. New York: Free Press.
- Cairney, P. (2012). Complexity theory in political science and public policy. *Political Studies Review* 10 (3): 346–58.
- Cairney, P. (2013). What is evolutionary theory and how does it inform policy studies? *Policy and Politics* 41 (2): 279–98.
- Campbell, D. (1965). Variation and selective retention in socio-cultural evolution. In H. Barringer, G. Blanksten, and R. Mack (Eds.), *Social Change in Developing Areas: A Reinterpretation of Evolutionary Theory*. Cambridge, MA: Schenkman, 19–49.
- Campbell, D. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The Philosophy of Karl R. Popper*. LaSalle, IL: Open Court, 412–63.
- Campbell, D. (1997). From evolutionary epistemology via selection theory to a sociology of scientific validity. *Evolution and Cognition* 3 (1): 5–38.
- Carstensen, M. B. (2011). Paradigm man vs the bricoleur: bricolage as an alternative vision of agency in ideational change. *European Political Science Review* 3 (1): 147–67.
- Carstensen, M. (2017). Institutional bricolage in times of crisis. *European Political Science Review* 9 (1): 139–60.
- Cashore, B. and Howlett, M. (2007). Punctuating which equilibrium? Understanding thermostatic policy dynamics in Pacific northwest forestry. *American Journal of Political Science* 51 (3): 532–51.
- Cilliers, P. (1998). *Complexity and Postmodernism: Understanding Complex Systems*. London: Routledge.
- Cohen, G. A. (1978). *Karl Marx's Theory of History: A Defence*. Oxford: Oxford University Press.
- Crouch, C. (2005). *Capitalist Diversity and Change*. Oxford: Oxford University Press.
- Crouch, C. and Farrell, H. (2004). Breaking the path of institutional development? Alternatives to the new determinism. *Rationality and Society* 16 (1): 5–43.
- Dewey, J. (1946 [1927]). *The Public and its Problems*. New York: Henry Holt.
- Dewey, J. (1948 [1922]). *Human Nature and Conduct*. New York: Henry Holt.
- Dopfer, K. (2012). The origins of meso economics. *Journal of Evolutionary Economics* 22 (1): 133–60.
- Dopfer, K. and Potts, J. (2014). *The New Evolutionary Economics*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Dowding, K. (2000). How not to use evolutionary theory in politics. *British Journal of Politics and International Relations* 2 (1): 72–80.
- Dronamraju, K. (2010). *Haldane, Mayr, and Beanbag Genetics*. Oxford: Oxford University Press.
- Dubash, N. and Morgan, B. (Eds.) (2013). *The Rise of the Regulatory State of the South*. Oxford: Oxford University Press.
- Eldredge, N. and Gould, S. (1972). Punctuated equilibria: an alternative to phyletic gradualism. In T. Schopf (Ed.), *Models in Paleobiology*. San Francisco, CA: Freeman, pp. 82–115.
- Elster, J. (1980). Cohen on Marx's theory of history. *Political Studies* 28 (1): 121–8.
- Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.
- Foster, J. and Pyka, A. (2014). Introduction: co-evolution and complex adaptive systems in evolutionary economics. *Journal of Evolutionary Economics* 24 (2): 205–7.
- Freeman, R. (2007). Epistemological bricolage: how practitioners make sense of learning. *Administration and Society* 39 (4): 476–96.
- Goertz, G. and Mahoney, J. (2012). *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.
- Gould, S. (2002). *The Structure of Evolutionary Theory* (5th edition). Cambridge, MA: Harvard University Press.
- Gronow, A. (2011). *From Habits to Social Structures: Pragmatism and Contemporary Social Theory*. Frankfurt am Main: Peter Lang.
- Hall, P. (1993). Policy paradigms, social learning, and the state. *Comparative Politics* 25 (3): 275–96.

- Haynes, P. (2008). Complexity theory and evaluation in public management. *Public Management Review* 10 (3): 401–19.
- Head, B. (2016). Toward more ‘evidence-informed’ policy making? *Public Administration Review* 76 (3): 472–84.
- Hodgson, G. (2002). *How Economics Forgot History: The Problem of Historical Specificity in Social Science*. London: Routledge.
- Hodgson, G. (2014). *Economics in the Shadows of Darwin and Marx*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Hoppe, R. (2002). Co-evolution of modes of governance and rationality: a diagnosis and research agenda. *Administrative Theory and Praxis* 24 (4): 763–80.
- Jabko, A. and Sheingate, A. (2018). Practices of dynamic order. *Perspective on Politics* 16 (2): 312–27.
- James, W. (1950 [1890]). *The Principles of Psychology*, 2 vols. New York: Dover Publications.
- Jervis, R. (1999). *System Effects: Complexity in Political and Social Life*. Princeton, NJ: Princeton University Press.
- John, P. (1999). Ideas and interests: agendas and implementation. *British Journal of Politics and International Relations* 1 (1): 39–62.
- John, P. (2000). The uses and abuse of evolutionary theory in political science. *British Journal of Politics and International Relations* 2 (1): 89–94.
- Kauffman, S. (1995). *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. Oxford: Oxford University Press.
- Kay, A. (2006). *Dynamics of Public Policy*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Kay, A. and Daugbjerg, C. (2015). De-institutionalising governance? Instrument diversity and feedback dynamics. *Asia Pacific Journal of Public Administration* 37 (4): 236–46.
- Kerr, P. (2002). Saved from extinction: evolutionary theorising, politics and the state. *British Journal of Politics and International Relations* 4 (2): 330–358.
- Kincaid, H. (1996). *Philosophical Foundations of the Social Sciences: Analyzing Controversies in Social Research*. Cambridge: Cambridge University Press.
- Kingdon, J. (1995). *Agendas, Alternatives and Public Policies* (2nd edition). New York: HarperCollins.
- Klamer, A. and Leonard, T. (1994). So what’s an economic metaphor? In P. Mirowski (Ed.), *Natural Images in Economic Thought*. Cambridge: Cambridge University Press, 20–52.
- Klijin, E. (2008). Complexity theory and public administration: what’s new? *Public Management Review* 10 (3): 299–317.
- Lakoff, G. and Johnson, M. (2003). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.
- Lewis, O. and Steinmo, S. (2010). Taking evolution seriously in political science. *Theory in Biosciences* 129 (2–3): 235–45.
- Lewis, P. (1996). Metaphor and critical realism. *Review of Social Economy* 54 (4): 487–506.
- Lustick, I. (2011a). Taking evolution seriously: historical institutionalism and evolutionary theory. *Polity* 43 (2): 179–209.
- Lustick, I. (2011b). Institutional rigidity and evolutionary theory: trapped on a local maximum. *Clodynamics* 2 (2): x–xx.
- Mahoney, J. and Thelen, K. (2010). A theory of gradual institutional change. In J. Mahoney and K. Thelen (Eds.), *Explaining Institutional Change: Ambiguity, Agency, and Power*. Cambridge: Cambridge University Press, 1–37.
- Marsh, D. and McConnell, A. (2010). Towards a framework for establishing policy success. *Public Administration* 88 (2): 564–83.
- Moran, M. (2002). Understanding the regulatory state. *British Journal of Political Science* 32 (2): 391–413.
- Moynihan, D. and Soss, J. (2014). Policy feedback and the politics of administration. *Public Administration Review* 74 (3): 320–332.
- Nelson, R. and Winter, S. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, MA: Belknap Press.
- Nungesser, F. (2017). The evolution of pragmatism: on the scientific background of the pragmatist conception of history, action, and sociality. *European Journal of Sociology* 58 (2): 327–67.

- Peters, B. G., Fontaine, G., and Mendez, J.-L. (2018). Substance and methods in the comparative study of policy change. *Journal of Comparative Policy Analysis* 20 (2): 133–141.
- Potts, J. (2000). *The New Evolutionary Microeconomics: Complexity, Competence and Adaptive Behaviour*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Potts, J. (2014). Innovation is a spontaneous order. *Cosmos and Taxis: Studies in Emergent Order and Organization* 2 (1): 1–10.
- Sabel, C. and Zeitlin, J. (2012). Experimentalist governance. In D. Levi-Faur (Ed.), *The Oxford Handbook of Governance*. Oxford: Oxford University Press, 169–84.
- Shpaizman, I. (2017). Identifying relevant cases of conversion and drift using the Comparative Agendas Project. *Policy Studies Journal* 45 (3): 490–509.
- Simon, H. (1978). Rationality as process and as product of thought. *American Economic Review* 68 (2): 1–16.
- Spicer, M. (2014). In defence of value pluralism in public administration. *Administration and Society* 46 (8): 1010–1019.
- Steinmo, S. (2010). *The Evolution of Modern States*. Cambridge: Cambridge University Press.
- Streeck, W. and Thelen, K. (2005). Introduction: institutional change in advanced political economies. In W. Streeck and K. Thelen (Eds.), *Beyond Continuity: Institutional Change in Advanced Political Economies*. Oxford: Oxford University Press, 1–39.
- Tenbenschel, T. (2013). Complexity in health and health care systems. *Social Science and Medicine* 93: 181–4.
- Tenbenschel, T. (2015). Complexity and health policy. In P. Cairney and R. Geyer (Eds.), *Handbook on Complexity and Public Policy*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing, 369–83.
- Tenbenschel, T. (2018). Bridging complexity theory and hierarchies: markets, networks, communities: a ‘population genetics’ framework for understanding institutional change from within. *Public Management Review* 20 (7): 1032–51.
- Thelen, K. (2004). *How Institutions Evolve: The Political Economy of Skills in Germany, Britain, the United States, and Japan*. New York: Cambridge University Press.
- van der Heijden, J. (2010). A short history of studying incremental institutional change. *Regulation and Governance* 4 (2): 230–243.
- Wagenaar, H. (1999). Value pluralism in public administration. *Administrative Theory and Praxis* 21 (4): 441–9.
- Weise, P. (1996). Evolution and self-organization. *Journal of Institutional and Theoretical Economics* 152 (4): 716–22.
- Whitford, J. (2002). Pragmatism and the untenable dualism of means and ends: why rational choice theory does not deserve paradigmatic privilege. *Theory and Society* 3: 325–63.
- Wilder, M. and Howlett, M. (2014). The politics of policy anomalies: bricolage and the hermeneutics of paradigms. *Critical Policy Studies* 8 (2): 183–202.

Index

- 21st Conference of the Parties Climate Convention (COP 21) 147
- Abel, T. 223
- ACA *see* Affordable Care Act (ACA)
- ACF *see* Advocacy Coalition Framework (ACF)
- actionable governance indicators 190, 192, 199
- Adecco Group 198
- adversarial subsystems 76
- Advocacy Coalition Framework (ACF) 67
- comparative research agenda 82–3
 - elements of comparison 68–70
 - guidelines, experiences and shortcomings 70–74
 - insights and lessons 74–5
 - conceptual classification 76–7
 - contextual description 75–6
 - hypothesis-testing 77–80
 - prediction 80–81
 - research program 68
- advocacy coalitions 67, 343
- Affordable Care Act (ACA) 353
- agency 395–6
- Agendas and Instability in American Politics* (1993) 92
- agenda-setting approach 101
- agenda-setting models 90–92
- agenda-setting research 102
- “Aims and Scope” statement 373
- Aklin, M. 178
- aligning process with realism 279–80
- Allardt, E. 24
- Allison, G. T. 161
- alternative coding strategies 213
- Alvarado, Velasco 122, 124, 126
- Amazonia in the Anthropocene* (Kawa) 321
- “ambassadors of participation” 144–5
- analysis of variance (ANOVA) paradigm 225
- Anand, N. 320
- Anckar, C. 45
- Andereggen, C. 262
- ANOVA paradigm *see* analysis of variance (ANOVA) paradigm
- anti-corruption policies 289
- artificial randomization 155
- Arts, B. 175
- Atkinson, M. M. 355
- atomist fallacy 245
- Avrami, S. 380
- Aykut, S. C. 147
- Bachrach, P. 90, 91, 102
- Baehler, K. J. 316–17
- Baratz, M. S. 90, 91, 102
- Baumgartner, F. 90, 92–4, 97, 99, 100, 102, 161, 169, 387
- Baumgartner, M. 268
- Bayes likelihood theorem 286, 291–2
- Bayesian designs 230–32
- Bayesian formalization 289
- of conventional tests 287
- Beach, D. 277
- Beath, A. 160
- “belief homophily hypothesis” 81
- Benito, B. 380
- Bennett, A. 74, 115, 277, 278
- Benton, A. 321
- Berejikian, J. 337
- Bergan, D. E. 161
- Berger, S. 359
- Berggren, J. 357
- Berg-Schlosser, D. 37, 38
- Berins-Collier, R. 116
- BESD *see* Binary Effect Size Display (BESD)
- Bevan, S. 95, 99
- Beyond Machiavelli: Policy Analysis Reaches Midlife* (Radin) 367
- Bhaskar, R. 220
- Binary Effect Size Display (BESD) 228
- biomedicalization of aging 300
- Biosafety Law 172
- Blair, B. D. 354
- Bond, J. R. 161
- Boon, J. 263
- bottom-up case-based approaches 243–5
 - see also* variance-based approaches
- Boydston, A. E. 99
- Boyle’s law 51, 53
- Brady, H. E. 115
- Brans, M. 262, 269, 322
- Bräuninger, T. 178
- Braybrooke, D. 395
- British thermal units (BTUs) 225
- Broockman, D. E. 161
- Brown, S. R. 335, 338, 339
- Brunner, R. 227
- Brusca, I. 380
- BTUs *see* British thermal units (BTUs)

- budget approach, for policy change 175–8
- Buell, E. 100
- Busch, P. 178
- Butler, D. M. 161
- Cacciatore, F. 265
- Calanni, J. 81
- Campbell, D. T. 27, 158, 220, 229, 390
- Cantú, F. 319, 320
- CAP *see* Comparative Agendas Project (CAP)
- Cappella, J. N. 306
- Cardoso, Fernando 171
- Cartwright, N. 240
- case-based approaches 238, 239, 243–5, 250, 251
- case-oriented approach 261–3
- causal case studies
 - bottom-up case-based approaches 243–5
 - of policy processes 245–50
 - top-down, variance-based approaches 240–43
- causal heterogeneity 239
- causal inference experiments 155–6
- causal process observations (CPOs) 115, 117
- ‘causes-of-effects’ approach 389
- Centre for Research and Expertise in Social Gerontology (CREGES) 300
- ceteris paribus* method 387
- Chaqués Bonafont, L. 100
- Checkel, J. 278
- Chin, M. L. 161
- circularity 396
- city-level indicators, measurement issues in 193–8
- Clinton, Hillary 350
- Cobb, R. 90–91, 95, 205
- Cohen, G. A. 389
- Coleman, W. D. 355
- collaborative subsystems 76
- Collier, D. 60, 115, 116
- Comparative Agendas Project (CAP) 90, 176, 203, 205, 213
 - agenda-setting models 90–92
 - measures of 97–100
 - methodological developments 92–5
 - policy agenda research 100–102
 - political agendas 95–7
- comparative case study design 228–30
- comparative historical analysis (CHA) 19, 113–15
 - assessing industrial policy 119
 - applying values to index concepts and indicators 122–4
 - checking internal validity 124–6
 - components and values of index 120–22
 - explaining causal mechanisms 126–9
 - index of policy intervention 117–19
 - maximizing process tracing results 115–17
- Comparative Manifesto Project (CMP) 93
- comparative method
 - and case studies 26
 - “evidence-based policy making” 20
 - modes of analysis 21–5
 - possibility of hybrid designs 26–7
 - sampling on dependent variable 25–6
 - small-N research 21, 28
- comparative policy analysis (CPA)
 - methodological issues 6–9
 - research design and methods selection 9–14
 - substance and methods in 1–3
 - theoretical issues 3–6
- comparative policy research and practice 14–15
- “comparative public administration” 1
- comparative toxics policy 320
- concept identification 210–11
- conceptual classification 74
- conceptual description 74
- “conceptual stretching” 348
- conceptualizing policy change 168–70
- condition-oriented approach 261–3
- Conditional Cash Transfers (CCT) 135
- Conducting Interpretive Policy Analysis* (Yanow) 315
- conflict management perspective 90, 91
- congruence analysis 288–9
- contacts associated with the research center (CREGES) 303
- contemporary public policy 4
- contemporary social sciences 28
- conventional comparative strategies 137
- Cook, T. D. 220
- Correa, Rafael 127, 128
- counterfactual causation 239
- CPA *see* comparative policy analysis (CPA)
- Cramer, Kathy J. 326
- critical juncture 116
- critical multiplism
 - Bayesian designs 230–32
 - bounding multiple triangulation 232–3
 - experimental binary effect size display 227–8
 - and mixed methods 220–21
 - meta-analysis and meta-regression analysis 223–5
 - multiple observers 223
 - multiple operationism 222–3
 - multiple theories and construct validation 223
 - multiple observer comparative and reflexive case study design 228–30
 - nested and crossed quantitative designs 225–6
 - nested qualitative designs 226–7
 - Qualitative Comparative Analysis (QCA) 232

- qualitative regression design 228
 - visual mixed-method designs 230
- critical realism 15
- crossed quantitative designs 225–6
- cross-sectional ACF research 70
- Crow, D. A. 357
- cultural policy 95
- cultural theory 358

- Dahl, R. 318
- Damonte, A. 255
- Darwinian evolutionary theory 390, 393
- De Graaf, G. 336, 338
- De Leon, J. 320
- De Meur, G. 37, 38
- de Wijs, L. 337
- DEC *see* Development Experience Clearinghouse (DEC)
- decision-making process 205, 314
- decompositional causal strategy 226
- deductive-nomological (D-N) explanation 221
- deductive research
 - most-different systems design in 37
 - most-similar systems design in 36–7
 - strategies 35–6
- deep within-case study 289–91
- Dente, B. 114
- Design of Experiments, The* (Fisher) 156
- Desmond, M. 326, 328
- Development Experience Clearinghouse (DEC) 233
- Dewey, J. 221
- Dickinson, H. 343
- Dicle, B. 381
- Dicle, M. F. 381
- diffusion approach, for policy change 178–80
- Dimitrova, A. 299
- “disaggregated and actionable indicators” 192
- discourse networks approach 205
- distinguishing statements 342
- Dodds, A. 355
- Dolowitz, D. 139
- DOS-based program 341
- Dossi, S. 114
- “doubly-engaged ethnography” framework 329
- Dowding, K. 395
- Down, Out, and Under Arrest* (Stuart) 319
- Dryzek, J. S. 337
- dualism 283
- Dunlop, C. A. 357

- EBM *see* Evidence-Based Medicine (EBM)
- EBSCO Academic Complete database 369, 371, 374–8, 380, 381
- ecological fallacy 242
- Ecuador
 - expanded time frame index 126
 - index of intervention 123
- Ecuadorian industrial policy 127
- Egami, N. 212
- E-Government Readiness Index 191
- Einstein, K. L. 162
 - relativity theory 54
- Elder, C. 90–91, 95, 205
- Eldredge, N. 387
- elite experiments 159–62
- Elster, J. 389
- “empirical counterfactual” elements 284
- empirical generalization 53–7, 61–3
- empirical realism 15
- empirical tests design 285–7
- Engeli, I. 270, 322
- Enhanced Standard Analysis (ESA) 266
- Enikolopov, R. 160
- Environmental Impact Assessment 172
- E-Participation Index 191
- epistemic fallacy 220, 222, 279
- “epistemic theory of causality” 282
- epistemological perspectives 303–5
- Ertas, N. 354
- ESA *see* Enhanced Standard Analysis (ESA)
- ethnographic engagement, three modes of 317–21
- ethnographic interviews 222
- ethnography 322–6
 - application of 313–15
 - promises and perils 326–8
 - defined 313
 - embracing 328–9
 - policy issue areas 323
 - in public policy/management/administration research 315–17
 - single-case 312
 - three modes 317–21
- European Innovation Scoreboard 194
- evidence-based approach 391
- Evidence-Based Medicine (EBM) 143
- evidence-based policy making 4, 20, 24, 234, 368, 395
- evolution
 - as metaphor 390–92
 - as process 386–9
 - as theory 386–90
- evolutionary biology 396
- evolutionary explanation 57, 395
- evolutionary frameworks, for institutionalist theory 392–4
- evolutionary theory 385, 395–6
 - agency 395
 - circularity 396
 - evolutionary approaches 386–92
 - functionalism 394–5

- intentionality 395
- experimental binary effect size display 227–8
- experiments 153–5
 - causal inference 155–6
 - elite 159–62
 - history in public policy 156–9
 - lessons for students of policy process 162–3
- exploratory approach 263
- export-oriented development policies 124

- Falleti, T. 113, 130
- feedback theory 228
- field experiment 155
- first-order policy change 168
- Fischer, M. 81, 262
- Fisher, R. A. 156, 157, 225
- Fisher, W. R. 349
- Flick, U. 220
- Flores, R. D. 206
- focus group methods 297–9
 - designing 305–6
 - managing group dynamics 307–8
 - methodological advantages 301–2
 - epistemological implications 303–5
 - revealing group consciousness 302–3
 - methodological disadvantages 305
 - in Québec, deployment of 300–301
- focus groups techniques 9
- Fontaine, G. 322
- formal set-theoretic theory evaluation 264
- Fotini, C. 160
- Foyer, J. 147
- Freeman, G. 24
- Fringe Special* initiative 191
- Fujimori, Alberto 124, 127
- functional explanation (FE) 389
- functionalism 394–5

- Galton's problem 137, 148, 225, 234
- Gapminder software 230
- Geddes, B. 25
- Geertman, S. 337
- genetically modified organisms (GMO) 171
- George, A. 74, 277, 278
- Gerber, A. S. 159
- Gerring, J. 54, 136, 138, 264, 302
- Gerrits, L. M. 265
- Geva-May, I. 315–16
- Geva, N. 161
- GG *see* Government at a Glance (GG)
- Gilardi, F. 205, 206
- Gleeson, D. 298
- Glick, D. M. 162
- Global Cities Talent Competitiveness Index 198
- Global Innovation Index 195
- Global Integrity Index 191
- Global Power City Index 195–6
- Global Talent Competitiveness Index 198
- GMO *see* genetically modified organisms (GMO)
- Goetz, J. P. 325
- Goffman, A. 319, 328
- Golder, B. 381
- González-Bailón, S. 203
- good governance 186, 188–95, 197, 199, 223
- Gould, S. 61, 387
- governance indicators and critique of ranking 189–90
- government budgets 176
- Government at a Glance (GG) 192
- gradualism 386
- Gray, G. 351
- Green, D. P. 159
- Greenberg, D. H. 158
- Green-Pedersen, C. 96, 99
- group consciousness 302–3
- group dynamics 307–8
- “group effect” 302

- Hackenberg 315
- Hale, S. A. 162
- Hall, P. A. 167–70, 182, 233, 388
- Halpern, C. 146
- Hanmer, M. J. 161
- Hassenteufel, P. 137, 143
- Hecló, H. 233
- Hedstrom, P. 58
- Heidenheimer, A. J. 355
- Henry, A. D. 355
- Hess, D. R. 161
- Heyse, L. 270
- HI *see* historical institutionalism (HI)
- Hinterleitner, M. 265
- historical institutionalism (HI) 114
- HIV Exceptionalism* (Benton) 321
- Hodgson, G. 390
- Hogwood, B. W. 167
- Hollander, J. A. 302
- Holzinger, K. 169–71, 175
- Homo narrans* model, of rationality 352–3, 361
- Hood River Energy Experiment 228
- Hooijer, G. 269
- Hoover, E. 320
- Houts, A. C. 220
- HRAF *see* Human Relations Area File (HRAF)
- Huberman, A. M. 221
- Huda, J. 359
- Huey, L. 319, 325
- Human Capital Leadership Institute of Singapore 198
- Human Relations Area File (HRAF) 224

- Humphreys, M. 230
Hydraulic City (Anand) 320
 hypothesis-testing 74
- IAT *see* Implementation Assessment Tool (IAT)
 imaginative reconstruction 223
 Implementation Assessment Tool (IAT) 192
 import substitution industrialization (ISI) 119
 indexes
 governance indicators and critique of ranking 189–90
 measurement issues in regional and city-level indicators 193–8
 second-generation measurements 190–93
 inductive research
 most-different systems design in 37
 most-similar systems design in 36–7
 strategies 35–6
 industrial policy 119
 applying values to index concepts and indicators 122–4
 checking internal validity 124–6
 components and values of index 120–22
 explaining causal mechanisms 126–9
 industrial protectionist laws 129
 inferential felony 25
 Ingold, K. 81
 Ingram, H. 169
 Innovation Cities Index 195
 institutional determinism 114
 institutionalist theory 392–4
 intentional selection mechanisms 395
 intentionality 395
 International Budget Partnership 191
 international development organizations 369
 international organizations 138, 144, 145, 147, 194, 379
 international policy transfers 135
 comparative analysis of 136–7
 comparative country analysis 137–8
 comparison in 140
 transfer agents, instruments, and arenas 144–8
 transnational processes 141–4
 issues, questions, and concepts for 138–40
 International Political Science Association 8
 International Public Policy Association 8
 invariant generalizations 51–8, 63
 ISI *see* import substitution industrialization (ISI)
- Jacobs, A. 230
 James, O. 162
 Jang, S. 81
 Jano, D. 262
- JCPA see* *Journal of Comparative Policy Analysis (JCPA)*
 Jeffares, S. 336
 Jervis, R. 388
 John, P. 91, 161, 162, 395
 Jones, B. 90, 92–4, 97, 99, 102, 161, 169, 387
 Jones, M. D. 348, 350, 351, 354
 Jörgens, H. 178
Journal of Comparative Policy Analysis (JCPA)
 315–16, 367, 374–7, 380
 countries studied 378–9
 policy areas studied 378
 publication trends 378
 ‘just so stories’ 61–2
- Kaplan, A. 373
 Kaplan-Meier failure function 178
 Kapur, D. 380
 Kassimeris, C. 381
 Kauffman, S. 387, 388, 391
 Kawa, N. C. 321
 Keohane, R. 49
 Kerr, P. 394
 King, G. 49, 56, 62
 Kingdon, J. 90, 91, 96, 387, 395, 396
 Klüver, H. 206
 Knieper, C. 265
 Knill, C. 169–71, 175
 Kortenska, E. 299
 Kübler, D. 71
- ladder of abstraction 335, 348, 356
 Ladi, S. 145
Land of Open Graves (De Leon) 320
 Landman, T. 80, 136, 138
 Lara, Rodríguez 124, 129
 large-N quantitative methods 226
 large-N regression analysis 219
 Lasswell, H. D. 373
 latent structures 212
 Laudan, L. 67
 Lebel, B. 356, 358
 Lebel, L. 356, 358
 LeCompte, M. D. 325
 Le Galès, P. 146
 lemmatization 210
 leptokurtosis 97
 Li, Q. 224
 Lieberman, E. S. 222, 225, 226
 Lijphart, A. 21, 33, 34, 36, 219
 Lin, P.-Y. 352
 Lindblom, C. 92, 368, 395
 Linder, F. 206
Line Becomes a River, The (Cantú) 319, 320
 lobbying 206

- Löblová, O. 249
 logical minimization process 256
 long-term opportunity structure 71
 Lora, E. 119
 Lora-Wainwright, Anna 321
 low-threshold solution 209
- Mackenzie, D. 221
 Maggetti, M. 254, 255, 260, 261, 263–7, 270
 Mahoney, C. 206
 Mahoney, J. 28, 60, 113, 130
Mano Dura (Wolf) 320
 Marsh, D. 139
 Mazmanian, D. 265
 McBeth, M. K. 348, 350
 McClendon, G. H. 161
 McClure, K. 361
 McCollum, Elmer 157
 McCormack, L. 354
 McGranahan, C. 313
 McKeown, B. 335, 338
 MCWC *see* minimum connected winning coalitions (MCWC)
 MDSD *see* most-different systems design (MDSD)
Measurement of Meaning, The (Osgood, Suci, and Tannenbaum) 224
 mechanistic evidence 244
 Mejía-Dugand, S. 141
 Mendez, J.-L. 322
 meta-analysis 223–5
 meta-regression analysis 223–5, 234
 methodological developments 92–5
 methodological individualism 3
 methodological nationalism 135, 136, 148
 Miles, M. B. 221
 Mill, J. S. 34, 37, 39, 41, 136, 368
 Millar, K. M. 321
 minimum connected winning coalitions (MCWC) 55
 Mitchell, A. 224
 mixed-gender focus groups 303
 mixed-methods
 meta-analysis and meta-regression analysis 223–5
 multiple observers 223
 multiple operationism 222–3
 multiple theories and construct validation 223
 research 220–21
 moderator involvement 306, 307
 Monroe, B. L. 203
 Montefrio, M. J. F. 381
 Montpetit, E. 322
 Moore, B. 114
 Morena, E. 147
 Morgan, D. L. 298, 306
 Mortensen, P. 96, 100
 Morton, R. B. 158
 Moseley, A. 162
 most-different systems design (MDSD) 23, 24, 34–5, 38–9, 41, 45
 in contemporary research 42–4
 in deductive and inductive research 37
 most-similar systems design (MSSD) 23, 33–4, 38–9, 44, 45, 238
 in contemporary research 42–4
 in deductive and inductive research 36–7
 interaction effects in 40
 technique 136
 MSF *see* multiple streams framework (MSF)
 MSSD *see* most-similar systems design (MSSD)
 multi-dimensional approach, for policy change 170–75
 multiple observer comparative case study design 228–30
 multiple operationism 222–3
 multiple streams framework (MSF) 265
 multiple triangulation 232–3
 multiple triangulation case study design 230
- Nagle, R. 321
 Named Entity Recognition (NER) 211
 narrative fidelity 361
 narrative persuasion 354, 359
 Narrative Policy Framework (NPF) 348–51, 355–7
 comparative applications 357–9
 limitations 360
 universality of narrative 351–2
 narrative and behavior (strategy and mobilization) 354–5
 narrative persuasion 354, 359
 narrative rationality 352–3, 361
 narrative probability 361
 narrative rationality 352–3, 361
 nascent and mature subsystems 76–7
 Natalini, A. 265
 National Reform Programmes 265
 necessity 11
 negative income tax 157
Negotiating Demands (Huey) 319
 Nelson, R. 390
 neoliberalism 369
 neo-positivist conception of causation 279
 neo-positivist orthodoxy 14
 NER *see* Named Entity Recognition (NER)
 nested analysis 222, 223, 225
 nested qualitative designs 225–7
 nested quantitative designs 225–6
 Neumayer, E. 39
 New Public Governance (NPG) 343

- Newman, I. 333, 335–6
 Nickerson, D. W. 161
 Nieto Morales, F. 270
 Nir, L. 306
 Nohrstedt, D. 75
 Nowlin, M. C. 205
 NPF *see* Narrative Policy Framework (NPF)
 NPG *see* New Public Governance (NPG)
 Nyhan, B. 161
- objective relativism 234
 O'Bryan, T. 357
 OECD *see* Organisation for Economic
 Co-operation and Development (OECD)
 O'Leary, R. 358
 Olofsson, K. 75
On the Run (Goffman) 319, 328
 one-on-one interviews 301
 ontological determinism 244
 Open Budget Index 191
 Open Net Initiative 192
 Oppen, M. 206
 Optical Character Recognition (OCR) software 208
 Orel, N. A. 301
 organic water pollutants
 emission standards 171
 stringency of limit values 173
 Organisation for Economic Co-operation and
 Development (OECD) 145–6
 Osgood, C. E. 224
 Osorio Gonnet, C. 142, 143
 Ostrom, E. 67
 outcome data, for policy change 180–82
 Owen, E. 224
 oxytocin 352
- Pacheco-Vega, R. 314, 328, 329
 Pahl-Wostl, C. 265
 Pal, L. A. 139, 146
 Palau, A. M. 100
 Parizeau, K. 314, 328, 329
 Park, K. 81
 parsimonious solution models (PS) 268
 part-of-speech (POS) 210
 “party system agenda” 96
 path-dependent trajectories 115
 Pattyn, V. 262, 269
 Pearl, J. 221
 Peck, J. 142
 Pedersen, R. B. 277
 Pelizzo, R. 299
 Peritore, A. G. 337
 Peritore, N. P. 337
 Peru 126, 127
 expanded time frame index 126
 index of intervention 123
 industrial policy incentives 125
 neoliberal reforms 128
 PET *see* punctuated equilibrium theory (PET)
 Peters, B. G. 137, 146, 147, 167, 168, 322, 355
Picking Up (Nagle) 321
 Picot, G. 269
 Pierce, J. J. 72, 75
 Pierson, P. 113, 115, 116, 124
 policy agendas 205
 research 100–102
 policy analysis, process tracing applications to
 278–9
 policy change
 budget approach for 175–8
 comparative examinations of 80
 conceptualizing 168–70
 diffusion approach for 178–80
 multi-dimensional approach for 170–75
 operationalization of 167
 outcome approach for 180–82
 policy diffusion 206
 policy feedbacks 206–7
 Policy Frames Project 206
 policy learning and change 69
 policy monopoly 92
 policy-oriented learning 79
 policy-oriented research 254
 political agendas 95–7
Politics of Information, The (Baumgartner and
 Jones) 99
 Popper, K. 34, 50
 Porto de Oliveira, O. 139
 PQmethod 338–9
 pragmatist concepts of habit 393
 Price, V. 306
 private-sector participation 127
 problem definition 205–6
 process theorization 283–5
 process tracing
 aligning process tracing with realism 279–80
 applications to policy analysis 278–9
 for policy learning 291–2
 realist process tracing as set-theoretic method
 280–81
 realist view on causal mechanisms 281–3
 in social sciences 277–8
 template for realist process tracing
 case selection 287–8
 congruence analysis 288–9
 deep within-case study 289–91
 empirical tests design 285–7
 process theorization 283–5
*Process Tracing: From Metaphor to Analytic
 Tool* (Bennett and Checkel) 278

- Process Tracing Methods: Foundations and Guidelines* (Beach and Pedersen) 277
- Productive Development Policies (PDPs) 121
- protectionist industrial law 122
- protectionist policy agendas 119
- Przeworski, A. 22, 24, 34, 35, 37, 38, 43–5
- PSAs *see* public service announcements (PSAs)
- psychometric tests 336
- Public Accountability Measures 192
- public agenda 96–7
- public policies 153, 373
- case-based approaches 243–5
 - history of experiments 156–9
 - variance-based approaches 240–43
- public policy analysis
- qualitative comparative analysis
 - advantages 256–9
 - limitations 259–60
 - usefulness 262–3
 - public policy model 136
 - public service announcements (PSAs) 352
 - punctuation equilibrium theory (PET) 91, 169, 170, 175, 176, 182
 - purposive (P) sample 339
 - push approach 317–18
- Q methodology 333
- execute and analyze 335–6
 - defining discourses 341–2
 - drawing conclusions based on discourses 342–3
 - gaining insight into concourse 336–7
 - P-sample 339
 - Q-sample 337–9
 - ranking by participants 340
 - philosophy 334–5
- QCA *see* qualitative comparative analysis (QCA)
- Q-sample 337–9
- Q-sorting process 340
- qualitative comparative analysis (QCA) 1, 25, 42–3, 232, 246, 277
- advantages 256–9
 - analysis 255–6
 - approaches to 260
 - assumptions of 255
 - case-oriented/condition-oriented approach 261–3
 - in comparative public policy analysis 256–60
 - limitations 259–60
 - substantively interpretable/redundancy-free models 265–70
 - as technique 254–6
 - theory-generating/theory-evaluating mode of reasoning 263–5
- qualitative experimental designs 380
- qualitative regression analysis 228
- qualitative regression design 228, 229
- qualitative semi-structured interviews 314
- quantitative-by-qualitative mixed-methods domain 221
- quasi-experimental approach 354
- Radaelli, C. M. 114, 357
- Radin, B. 322, 367, 368
- Ragin, C. 42, 70
- Ramlo, S. E. 334–6
- randomized controlled trial 155
- rational institutionalism 113
- realism 279–80
- realist process tracing 282, 292
- case selection 287–8
 - congruence analysis 288–9
 - deep within-case study 289–91
 - empirical tests design 285–7
 - process theorization 283–5
- realistic evaluation approaches 259
- Reclaiming the Discarded* (Millar) 321
- reductionism 387
- reflexive case study design 228–30
- regional-level indicators, measurement issues in 193–8
- regression equations 223
- regularity 11
- regulation organic water pollution 179
- Reifler, J. 161
- Reno, J. O. 321
- Richardson, L. 162
- Riggs, F. W. 369
- Rihoux, B. 270
- Rimmerman, A. 380
- Robinson, J. 138
- Rook, D. W. 303
- Rose, R. 134, 368, 373, 374
- Rothmayr Allison, C. 270, 322
- Runhardt, R. W. 242
- Ryan, M. 161
- Sabatier, P. 71, 265
- Sager, F. 262, 265
- Sartori, G. 24, 35, 348, 351, 355
- Schattschneider, E. E. 90–91
- Schmidt, V. A. 353
- Schneider, A. 169
- Schneider, C. Q. 264, 266
- Schoonvelde, M. 209
- Schrodt, P. A. 203
- Schulze, K. 169–71, 175
- Schumacher, G. 209
- scientific knowledge 279
- Seawright, J. 115

- second-generation governance indicators 186–8, 190, 192, 193, 199
- second-generation measurements 190–93
- second-order policy change 168
- semi-structured interviews 329
- Shadish Jr, W. R. 220
- Shadish, W. R. 220
- Shahidi, F. V. 265
- Shamdasani, P. N. 303
- Shanahan, E. A. 348, 350, 357
- Sheffer, L. 161
- Shipan, C. R. 205, 206
- Shpaizman, I. 391
- Sigelman, L. 100
- Simon, H. 91, 92, 161, 395
- Skelcher, C. 336
- Skocpol, T. 35, 114, 230, 328
- small-N research 21, 28, 226
- Smith, D. 206, 354
- Social Origins of Dictatorship and Democracy* (Moore) 114
- social sciences, methodological shift in 6–7
- Song, G. 354
- Spearman, Charles 335
- Spencer, H. 389
- standard ordinary least squares regression 35
- Stapenhurst, R. 299
- States and Social Revolutions: A Comparative Analysis of France, Russia and China* (Skocpol) 35, 114, 230
- Stenner, P. 335, 340–41
- Stephenson, William 333–6, 339, 340
- Stewart, D. W. 303
- Stone, D. 136, 147, 148, 349
- Stramp, N. 206
- stratified ontology 9, 10
- structural topic model 205, 212
- structure-versus-agency debate 3
- structured interviews 329
- structuring principle 340
- Stuart, F. D. 319
- substantive policy domains 7
- Suci, G. J. 224
- Sugiyama, M. S. 352
- sui generis* methodological techniques 137
- supervised classification tasks 211
- Swanson, G. 371
- symmetry 11
- System of Logic, A* (Mill) 368
- Tannenbaum, P. H. 224
- taoism 42
- tariff protection scheme 120, 121
- tests of theory 49–51
 - empirical generalization 51, 54–6
 - invariant generalization 51–4
 - mechanism 51, 56–9
- Teune, H. 22, 24, 34, 35, 37, 38, 43–5
- text corpus 208–9
- text-as-data methods 203
 - applications 204–5
 - lobbying 206
 - new directions of 212–13
 - policy agendas 205
 - policy diffusion 206
 - policy feedbacks 206–7
 - problem definition 205–6
 - methods
 - classification 211
 - concept identification 210–11
 - discovery of latent structures 212
 - in practice 207
 - preparing text corpus 208–9
 - preprocessing 209–10
- Theodore, N. 142
- theory 49–51
 - testing of
 - empirical generalization 51, 54–6
 - invariant generalization 51–4
 - mechanism 51, 56–9
- theory-driven observation 265
- theory-driven process tracing 277
- theory-evaluating approach 263–4
- theory-evaluating mode of reasoning 263–5
- theory-generating mode of reasoning 263–5
- Thiem, A. 268
- Third-order policy change 168
- Thomann, E. 254, 255, 260–67, 270
- Thomas, D. 335, 338
- Thomas, H. F. 99
- Thomas, M. A. 223
- time frame validation 113
- time-series analysis 6
- top-down, variance-based approaches 240–43
- Tosun, J. 169–71, 175
- traditional deductive hypothesis testing 264
- traditional theory-generating approach 263
- transcendental realism 279
- transfactualism 283
- transnational processes 141–4
- transnationalization of public action 137
- Tummers, L. 263
- Udy, Jr, S. H. 224
- unique cases 60–62
- unit homogeneity 241, 242
- unitary policy subsystems 76
- United Nations e-Government survey 191
- universality of narrative 351–2

- narrative and behavior (strategy and mobilization) 354–5
- narrative persuasion 354
- narrative rationality 352–3
- Urpelainen, J. 178

- valve theory 228
- van Engen, N. 263
- Van Evera, S. 54
- Van Exel, J. 336, 338
- van der Heijden, J. 262
- van Hulst, M. J. 314, 324
- Van de Walle, S. 322
- Vannoni, M. 161
- variance-based approaches 239–43
- Varma, R. 380
- Varone, F. 81
- Velasco, F. 127
- Verba, S. 49
- Verhoest, K. 263
- Verweij, S. 265
- visual mixed-method designs 230
- Vleugel, Roger 191
- Vogtz, F. 316, 317
- Vries, E. de 209
- Vugteveen, P. 343

- Wagemann, C. 264, 265
- Waldner, D. 61
- Walker, J. L. 170, 206
- Washington consensus 190, 193, 369
- Waste Away* (Reno) 321
- Watts, S. 335, 340–41
- Weber, M. 226

- Weible, C. M. 71, 76, 81, 356
- Weimer, D. L. 322
- Weise, P. 388
- welfare-to-work policy 153–4
- Whewell, W. 234
- White, H. 245
- Who's Who (2008) (Thomson Reuters) 367
- Wildavsky, A. 92, 368
- Wilder, M. 355
- Wiles, J. L. 301
- Wilkerson, J. 206
- Wilkinson, S. 304
- Williams, G. 381
- Williams, K. C. 158
- Winter, S. 390
- Witte, P. 337
- Wittek, R. 270
- Wolf, A. 316–17
- Wolf, S. 320
- Wood, A. 142
- Woodward, J. 58
- Wordfish method 204, 206, 212
- Worldwide Governance Indicators (WGI) 189, 190, 192, 193, 195, 196, 199, 223
- Worthy, J. 161
- Wüest, B. 205, 206
- Wutich, A. 301
- Wylie, S. A. 320–21

- Yanow, D. 315
- Ylikoski, P. 58

- “zoom in and zoom out” strategies 147