OXFORD

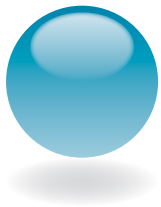
POLITICAL RESEARCH

METHODS AND PRACTICAL SKILLS

3RD EDITION

SANDRA HALPERIN | OLIVER HEATH

Political Research: Methods and Practical Skills



Political Research

Methods and Practical Skills

THIRD EDITION

Sandra Halperin and
Oliver Heath

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© Sandra Halperin & Oliver Heath 2020

The moral rights of the authors have been asserted

First Edition 2012
Second Edition 2017

Impression: 1

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, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2019951496

ISBN 978–0–19–882062–8

Printed in Great Britain by
Bell & Bain Ltd., Glasgow

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

Brief Contents

Detailed Contents vii
How to Use this Book xvii
Guided Tour of the Online Resources xviii
Preface xxi

1 Political Research 1

PART 1: **Philosophy of Social Science: Knowledge and Knowing in Social Science Research**

2 Forms of Knowledge: Laws, Explanation, and Interpretation in the Study of the Social World 27

3 Objectivity and Values 64

PART 2: **How to do Research: An Overview**

4 Asking Questions: How to Find and Formulate Research Questions 93

5 Finding Answers: Theories and How to Apply Them 123

6 Research Design 159

7 What is Data? 186

PART 3: **How to do Research in Practice**

8 Experimental Research 211

9 Comparative Research 231

10 Historical Research 260

11 Surveys 284

12 Interviewing and Focus Groups 309

13 Ethnography and Participant Observation 339

14 *Textual Analysis 364

15 Quantitative Analysis: Description and Inference 393
16 Patterns of Association: Bivariate Analysis 422
17 A Guide to Multivariate Analysis 451

Appendix 481
Glossary 487
Index 497

Detailed Contents

How to Use this Book xvii

Guided Tour of the Online Resources xviii

Preface xxi

1 Political Research 1

Issues in political research 2
Politics and International Relations 3
Empirical vs normative research 4
Positivism vs interpretivism 5
Quantitative vs qualitative research 6
The research process 7
Part 1: Philosophy of Social Science: Knowledge and Knowing in Social Science Research 8
Part 2: How to do Research: An Overview 10
Part 3: How to do Research in Practice 14
Conclusions 22
References 22

PART 1: **Philosophy of Social Science: Knowledge and Knowing in Social Science Research**

2 Forms of Knowledge: Laws, Explanation, and Interpretation in the Study of the Social World 27

Chapter Summary 27
Introduction 27
Positivism 29
Classical positivism 30
Empiricism and logic as the basis of truth claims 33
Challenges to positivist approaches in the social sciences 40
Scientific realism 40
1. What is the ontological status of macro-social mechanisms used to explain social outcomes? 42
2. How do we explain macro-social phenomena? 43
3. How do macro-social 'social mechanisms' produce social outcomes? Providing explanations of macro-social phenomena with micro-foundations 44
Interpretivism 47
The analysis of ethnic conflict: a positivist (rational choice), interpretivist (constructivist), and critical realist approach 51
A critical realist approach to explaining the Yugoslav conflict 57

Conclusions 59
Questions 61
Guide to Further Reading 61
References 62
Endnotes 63

3 Objectivity and Values 64
Chapter Summary 64
Introduction 64
Normative and empirical theory in political research 66
Values, the researcher, and the research process 67
Values and social practice 73
Thomas Kuhn and scientific revolutions 73
Paradigms and paradigm change 74
Kuhnian paradigms and political studies: the case of development theory 77
Paradigms and the study of development 78
Imre Lakatos and scientific research programmes 82
The role of values in research: Kuhn and Lakatos compared 86
Conclusions 87
Questions 88
Guide to Further Reading 88
References 89

PART 2: **How to Do Research: An Overview**

4 Asking Questions: How to Find and Formulate Research Questions 93
Chapter Summary 93
Introduction 93
Research questions: what are they? Where do they come from? 97
Finding research questions 97
How to formulate a research(able) question 104
The research vase 104
Types of questions: getting clear about what you want to know 107
Descriptive questions 108
Explanatory questions 109
Predictive questions 110
Prescriptive questions 110
Normative questions 111
Unanswerable questions: fallacies in framing research questions 112
'Researchability': a final consideration 114
The literature review 114
Stage 1: reading the literature 116
Stage 2: summarizing the literature 117
Stage 3: what still needs to be done? Setting the stage for your own argument 118

Conclusions 120
Questions 121
Guide to Further Reading 121
References 122

5 Finding Answers: Theories and How to Apply Them 123
Chapter Summary 123
Introduction 123
Answers to research questions: general requirements 127
What type of answer does your question require? 127
Answers that contribute to the development of theory 130
Where do hypotheses come from? 132
Illustration: how an analysis of existing theoretical perspectives provides the basis for a hypothesis 134
Specifying your argument or answer: the nuts and bolts 139
The components of a hypothesis 140
Variables 140
Relationships 142
Conceptualization and operationalization 143
Concepts and concept formation: what are you talking about? 144
Concept formation 144
Conceptualizing 'power': the search for theoretical utility 145
Conceptualizing 'democracy': mediating between coherence and differentiation 147
Operational definitions: how will you know it when you see it? 149
Illustration: Lenin's explanation of the First World War 150
Answering normative questions 152
Empirical and normative research 152
Empirical and normative questions, answers, and methods 154
Conclusions 155
Questions 156
Guide to Further Reading 156
References 157

6 Research Design 159
Chapter Summary 159
Introduction 159
Basic principles of research design 161
Types of research design 163
Experimental designs 163
Cross-sectional and longitudinal designs 165
Comparative designs 167
Historical designs 168
Types of research, types of design 170
Data-gathering strategies: how the data are collected 171
Questionnaires and surveys 172

Interviewing and focus groups 173
Focus groups 173
Ethnographic research 174
Discourse/content analysis 174
Ethical research 175
Ethical issues to consider concerning research participants 177
Informed consent 177
Maintaining confidentiality 179
The possibility of causing harm to participants 179
Ethical issues to consider relating to the researcher 180
Avoiding bias 180
Incorrect reporting 180
Inappropriate use of the information 180
Ethical issues regarding the sponsoring organization 180
Restrictions imposed by the sponsoring organization 180
The misuse of information 180
Conclusions 181
Questions 182
Guide to Further Reading 182
Ethical Research: Additional Resources and Readings 183
References 184
Endnotes 185

7 What is Data? 186
Chapter Summary 186
Introduction 186
From concepts to measurement 187
Validity 189
Example: measuring corruption 190
Reliability 191
How to choose cases 192
Types of data 195
Quantitative and qualitative data 196
Levels of measurement 197
Interview data 199
Observational data 199
Documents 200
Secondary data 200
Big data 201
Data quality checklist 204
Conclusion 205
Questions 205
Guide to Further Reading 206
References 206

PART 3: **How to do Research in Practice**

8 Experimental Research 211
Chapter Summary 211
Introduction 211
Basic principles of experimental design 212
Laboratory experiments 214
Field experiments 218
Natural experiments 222
Issues in experimental research 225
Conclusions 227
Questions 228
Guide to Further Reading 228
References 229

9 Comparative Research 231
Chapter Summary 231
Introduction 231
The principles of comparison 232
How we compare 234
Case study 234
Small-N comparison 237
Most Similar Systems Design 239
Example 1: the political salience of ethnic cleavages 239
Most Different Systems Design 241
Example 2: revolutions 242
Issues in small-N research 243
Qualitative comparative analysis (intermediate N) 247
Quantitative analysis (large-N comparison) 251
Conclusions 255
Questions 255
Guide to Further Reading 256
References 256

10 Historical Research 260
Chapter Summary 260
Introduction 260
Historical approaches in Politics and International Relations 261
Comparative and historical research 262
The timing and sequence of political events 265
Historical institutionalism 266
Process tracing 269
Event structure analysis 270
Using documents, archival sources, and historical writing as data 275
Identifying sources of data in historical research 275
Where to locate sources 276

Primary sources 276
Secondary sources 276
Evaluating sources 277
Primary sources 277
Secondary sources 278
Conclusion 279
Questions 280
Guide to Further Reading 280
References 281
Endnote 283

11 Surveys 284
Chapter Summary 284
Introduction 284
The elements of a good survey 285
What you ask: questionnaire design 286
Validity: from concepts to survey questions 287
Measurement error: developing indicators 288
Question wording comprehension 288
Recall problems 290
Deliberate misreporting 291
Question order effects 291
Question wording effects 293
Balanced questions 295
Response format 296
Who you ask: sample design 296
Defining the population: sampling frame 297
Choosing the sample: random probability sampling 297
Non-probability sampling 300
Sampling error 301
How you ask: interview mode 302
Conclusions 306
Questions 307
Guide to Further Reading 307
References 308

12 Interviewing and Focus Groups 309
Chapter Summary 309
Introduction 309
Interviews 310
Types of interview 310
Individual face-to-face interviewing 310
Telephone interviews 311
Online interviews 311

Focus groups 311
Advantages and disadvantages of different types of interview 311
Forms of interview 312
Unstructured interviews 313
Semi-structured interviews 313
Which type and form of interview to use? 313
Designing an interview guide for semi-structured or unstructured interviews 315
Conducting a face-to-face interview 318
Interviewing skills 318
Recording the interview 318
Keeping control over the interview 318
Probing 319
Ending the interview 319
Conducting online interviews 319
Email interviews 320
Interviewing via video link and web camera 321
Elite interviews 322
Expert interviews and surveys 324
Focus groups 325
Organizing and conducting a focus group session 327
Analysing interview data 329
Data reduction 329
Coding 330
Analysis 331
Conclusions 333
Questions 335
Guide to Further Reading 336
Software 337
References 337
Endnote 338

13 Ethnography and Participant Observation 339
Chapter Summary 339
Introduction 339
The principles of ethnography and participant observation 340
Theory and hypothesis testing 343
Participant observation methodology 344
Choosing a research site 346
Types of research site 346
Selection of research site 347
Access to research site 349
The role of the ethnographer 350
Doing participant observation 352
Choosing informants 352

Studying with informants 354
Identity of the researcher 355
Recording observations 357
Field notes 357
Conclusions 359
Questions 360
Guide to Further Reading 360
References 361

14 Textual Analysis 364
Chapter Summary 364
Introduction 364
Discourse analysis 365
Speech act theory, post-structuralism, and critical discourse analysis 366
Speech act theory 366
Post-structuralism 367
Critical discourse analysis 367
Analysis 368
1. The context 369
2. Analysing the process through which the power of a discourse has demonstrable effects 371
Validity and reliability 373
Content analysis 373
Qualitative and quantitative content analysis 376
Steps to carrying out content analysis 377
Step one 377
Step two 378
Step three 378
Step four 379
Coding 380
Manual and computer-assisted coding 380
Analysis 383
Validity and reliability 384
Conclusions 386
Questions 388
Guide to Further Reading 388
References 390
Endnotes 392

15 Quantitative Analysis: Description and Inference 393
Chapter Summary 393
Introduction 393
Descriptive and inferential statistics 394
Summarizing the data 395
Tables 396
Graphs and figures 400

Summarizing the distribution using statistics 402
Measures of central tendency 402
Measures of dispersion 405
Descriptive inference 410
Population estimates 412
Confidence intervals for proportions 412
Confidence intervals for the difference between two proportions 414
Confidence intervals for means 417
Conclusions 419
Questions 419
Guide to Further Reading 420
References 420

16 Patterns of Association: Bivariate Analysis 422
Chapter Summary 422
Introduction 422
The principles of bivariate analysis 423
Is there a relationship between the variables? 424
Null findings 425
How can we describe the relationship between the variables? 425
How good is our explanation? 426
Association and causality 426
Data considerations for bivariate analysis 426
Bivariate analysis for interval data 427
Bivariate analysis using scatter plots: describing the relationship 427
Is there a relationship? Significance testing with interval variables 431
The test statistic 433
The strength of the relationship: a question of fit 434
Pearson's correlation 436
Example 1: democracy and development 437
Bivariate analysis for categorical data: cross-tabulation 438
Bivariate analysis using tables: describing the relationship 439
Is there a relationship? Statistical significance for cross-tabulations 441
Example 2: ethnic diversity and civil war 446
Conclusions 448
Questions 448
Guide to Further Reading 449
References 449

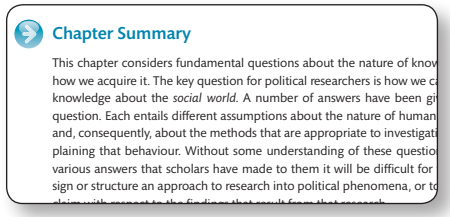
17 A Guide to Multivariate Analysis 451
Chapter Summary 451
Introduction 451
The principles of multivariate analysis: statistical control 452
Specifying different types of relationship 454
Spurious relationships 455

Intervening variables and indirect causality 456
Moderator variables and interaction effects 457
Multivariate analysis using OLS regression 458
Example 1: democracy 458
Measures of fit 463
Interpreting the magnitude of the coefficients 463
Presenting the results 465
Multivariate analysis using logistic regression 466
Example 2: turnout 469
Measures of fit 472
Interpreting the magnitude of coefficients 472
Interaction effects 474
Presenting and reporting data 478
Conclusions 478
Questions 479
Guide to Further Reading 479
References 480

Appendix 481
Glossary 487
Index 497

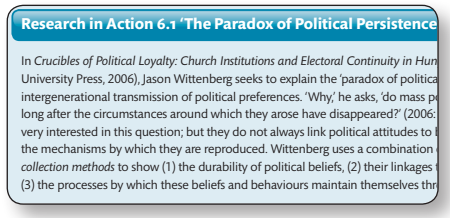
How to Use this Book

This book is enriched with a number of learning tools to help you to reinforce your knowledge, further your understanding, and develop your ability to think critically.



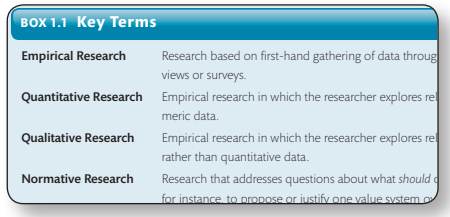
Chapter Summaries

Chapter overviews set the scene for upcoming themes and issues to be discussed, and indicate the scope of coverage within the chapter.



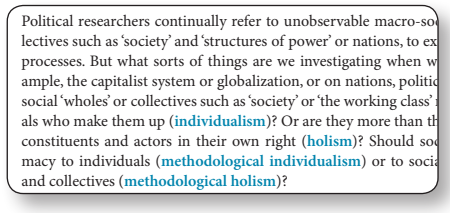
Research in Action Boxes

Research in action boxes provide summaries of published research to help you put your research project into practical context.



Boxes

Throughout the book, boxes provide you with extra information on particular topics and define and explain key ideas, helping you to build on what you have learnt from the main chapter text.



Glossary Terms

Key terms appear emboldened in blue in the text and are defined in a glossary at the end of the book, helping you to understand technical terms as you learn, and acting as a useful prompt when it comes to conducting your politics research project.

Guided Tour of the Online Resources

www.oup.com/he/halperin3e

The Online Resources that accompany this book provides students and instructors with a range of ready to use teaching and learning resources.

For Students

Examples of Research

Links to examples of different types of research help to you to evaluable key political research findings.

Links to datasets

Links to publically available datasets for Politics and IR research provide extra support for conducting your research project.

Web links

Relevant web links to reliable online content and databases related to topics in covered in the book are provided in the Online Resources for easy access.

For Registered Lecturers

Seminar Exercises

Suggestions for group activities to be carried out in seminars help support students understanding of the research project process.

Seminar Questions

Reinforce key themes from each chapter with suggested seminar questions.

Preface

Students devote a lot of time and effort to working on undergraduate, MA, and PhD theses, and there are a large number of texts available which can help with various aspects of this task. Some texts are devoted to the various debates and controversies relating to the philosophy of the social sciences, others to specific aspects of the research process, or to different research methods and approaches. But sometimes the various aspects of research that are the focus of different texts can seem disconnected, not only from each other, but from the process of pursuing knowledge about the political world. What we have tried to do is to incorporate all these different aspects of research within a single text and to discuss, step by step, how all of them relate to the research process. We begin with (1) an overview of key issues in the philosophy of social science (the problems of what we can know and how we can know it, in politics and in general). We then present (2) a 'nuts and bolts' or '*how to*' of research design: *how to* find and formulate a research question; *how to* engage with literature to develop a rationale for both the question and the approach that the researcher proposes to take in answering it; *how to* construct a theoretical framework and define concepts that provide help in answering the question. Then, in the final part of the book, we examine (3) different methods of data collection and analysis that can be used to answer research questions, the principles and procedures that guide the employment of each, and the variety of considerations and decisions that researchers must confront when using different methods.

Our aim is to provide students with an understanding of the perspectives, assumptions, logics, and methods that contribute to political research, as well as an accessible step-by-step guide through the different phases of a research project. We hope that this will help you to learn how to read analytically and think systematically about issues, problems, puzzles, and questions relating to the political world; and how to design and conduct independent research.

This book is based in large part on our experience of teaching at undergraduate and graduate levels. We would like to thank our students, whose questions, queries, and problems have informed many aspects of this book, and who have offered comments and criticisms on early drafts. In particular, we would like to thank Asma Ali Farah, Didem Buhari Gulmez, Baris Gulmez, Daniela Lai, and Ellen Watts for their help with the online resources for this text. Finally, we would like to thank our editor for this third edition, Katie Staal, and all the team at OUP who worked on this book, for their support, and the anonymous reviewers for their valuable feedback.

Political Research

This book has been written for undergraduate and graduate students of politics, with two main purposes in mind. These are to provide you, first, with the analytic skills and resources you need to evaluate research findings in political research; and, second, with the practical skills you need to carry out your own independent research. These twin goals are important for getting the most out of your study of politics.

The study of politics can often be complicated and confusing. In studying politics, we are frequently confronted with opposing ideas and arguments about a wide variety of different political phenomena. Is multiculturalism doomed to failure? Has globalization undermined national sovereignty? Is there a crisis of democracy and participation? Is conflict an inevitable consequence of religious, ethnic, and social difference? The answers to these questions, whether provided by academics, politicians, or journalists, can be inconsistent and contradictory. How do we know what to believe or who to trust?

A solid training in research skills can help us make sense of conflicting arguments and interpretations—to distinguish between those that are relatively sound and robust, and those that are unsubstantiated or rely on misleading or faulty inference. These skills are therefore crucial for helping us to make sense of the world. They help us to evaluate the merits of different arguments and the research of others, and to make our own arguments strong and convincing. Learning research skills is an active process that engages you in developing the ability to investigate the world around you and discover things for yourself. Pursuing research that enables you to find your own answers to questions, rather than just relying on what is said or has been written by others, can be exciting and challenging. It can lead you into new and surprising terrain.

These skills are at the core of political research. And understanding them and being able to use them transforms you from being a passive recipient of knowledge into an active protagonist. As students of politics, you are not only acquiring knowledge about the world of politics, you are also joined to a research community. Through engagement with research and writing in our field, and the independent thought and research you pursue in your own research projects, dissertations, or theses, you contribute to knowledge about the political world. But these skills also have wider relevance. They enable you to find creative solutions to problems and hone your analytical skills. They are genuinely transferable skills in that they can be used to sharpen and expand your thinking in whatever you do. We hope that through reading this book, and by engaging seriously with the principles and practices of political research, you will not only be more informed or knowledgeable about political affairs, but also become engaged, yourself, in a search for solutions to important problems of a public, political, or collective nature.

This book focuses on how to ask and answer questions about the political world. The ways in which we go about this, the methods that enable us to conduct systematic inquiry, allow us to connect our ideas about the way the political world works to evidence of what actually happens. Students sometimes view the study of methods as a kind of imposition—a set of rules and constraints designed to mould research into conformity with the conventions of a given field of study. This is misleading! Methods are not constraining: they are enabling.

Research methods are essentially about how to make arguments. All of us already *know* how to do this. We make them every day. We come up with well-reasoned arguments why others should believe what we believe, or why our way of doing something is better than other ways. And to make our arguments convincing we sometimes illustrate them with examples. Thus, research methods are embedded in the ways that we normally think and reason about things. What social science research requires you to do is to apply the skills of reasoning and argumentation you use in everyday life to larger questions of political life, and to hone these skills by thinking about what sort of evidence or examples you need to really support your argument. We have all learned to be wary about the use of anecdotal evidence. We don't consider the friend of a friend who saw or heard something to be a reliable source, and will often discount what people say if it is based on unreliable evidence. Research skills simply build on these intuitive everyday skills that we employ all the time. They are an organic and creative aspect of thinking and problem-solving. Moreover, they are intrinsically linked to the substantive concerns of our field. In our view, devising a research strategy that enables you to investigate or demonstrate an argument, hunch, or hypothesis is one of the really creative aspects of doing political research. It is the aspect of research that perhaps provides the greatest scope for independent thinking and creativity. Methods help you to build upon or synthesize the work of others, to connect up the work of different writers and thinkers with each other, or link together separate areas of study or studies of a single issue, in a way that generates fresh insights, that expands, extends, refines our knowledge of a political problem, puzzle, issue, system, process, structure, issue, or event.

Our approach to political research consists of three basic positions. *First*, we encourage pluralism in methodological approaches to political research. Politics is a multi-method field of study. The authors of this text reflect this pluralism. One of us pursues qualitative and exploratory historical-comparative work; the other quantitative and comparative work. *Second*, we believe that research should be *problem*-driven: that a research method only makes sense in so far as it is a way of investigating some problem. The value of any method can only be determined in relation to a research question; and the choice of which to use should always be driven by the research question that is to be investigated (see, for background and debate on this issue, Laitin 2005 and Shapiro 2005a, 2005b). *Third*, we believe research should have relevance to important political questions and policy issues. For research to be worth pursuing, we have to be interested in the question, and interested in the answer.

Issues in political research

This book, then, is concerned with how to formulate questions that are significant and how to develop meaningful and persuasive answers to them. A *significant question* is one that is 'directly relevant to solving real-world problems and to furthering the goals of a specific

scientific literature' (King, Keohane, and Verba 1994: 18). There is not always agreement about what constitutes a *meaningful and plausible answer* to a research question. But we can all agree that our answers should help to generate valid and reliable knowledge about the questions that they address. This requires that answers be developed through a process of inquiry that, at every step, is both self-aware and critical, and that researchers make clear and transparent how their conclusions were reached.

What research practices and methods enable political researchers to conduct systematic inquiry and thus make it possible for them to offer credible answers to important questions? This question has generated considerable debate.

Agreement on this question is difficult to achieve, in part because politics is such a diverse discipline. It struggles to even agree on a name. The study of politics is carried on in Departments of Politics, or Government, or Political Science or, in the UK, Departments of Politics and International Relations, to name just a few. The array of names reflects a variety of views about what constitutes the discipline. The growing tendency for sub-fields of political research, such as International Relations, to become institutionalized as almost separate fields of inquiry further fragments the field of study. Other divisions exist within the discipline, as well. There is a tendency to define **empirical** issues and the study of the 'real world', and normative issues and the study of political 'ideas and values', as involving separate research areas and traditions. And even scholars who study 'real-world' problems are divided about what constitutes empirical knowledge and how it can be achieved. Although these divisions are the focus of lively debate and disagreement within the discipline, they should not be overstated. Ultimately many of these oppositions are not as profound as they are often presented as being, and there is still more that unites the discipline than divides it. In the rest of this section we discuss some of these controversies in more detail.

Politics and International Relations

In recent years there has been a tendency to treat Politics and International Relations as separate fields of study, each with their own professional associations, journals, degree schemes, and even, in some universities, different departments. The establishment of these sub-fields institutionalized a division of the political world into processes and structures that are *internal* to states (local, domestic, or national politics), and those that are *external* to them (interactions and relations among states). However, like many in our field, we have come to question whether, within the general study of politics, it is productive to treat domestic and international processes, systems, events, and issues as analytically distinct. Increasingly, scholars in our field are recognizing that this distinction often obfuscates fundamental interdependencies of interstate and domestic systems. Politics and International Relations are both concerned with states, political systems, development, conflict, ideology, social movements, geopolitics, nationalism, political participation, and political philosophy. All aspects of politics are affected by governments, public administration and public policy, elections and voter behaviour, political parties, political culture, mass publics and electorates, interest groups, and public opinion, as well as by the interactions among states, the structure of the international system, international political economy, international organizations, and international law.

Researchers have increasingly come to appreciate the extent to which political processes operate across levels, or at multiple scales, of analysis. Societies interact, not just with their own governments, but with other societies and governments, and with international organizations; and because states have to answer to groups within their societies, a consideration of domestic relations is necessary to explaining outcomes in international relations. With a growing awareness of structural linkages between societies and the multiple channels that connect societies has come a tendency not to assume the sufficiency of any one scale of analysis to an understanding of political outcomes. Consequently, this book addresses the study of Politics and of International Relations as a single area of inquiry; and we address ourselves, as well, to students of all sub-fields of political research, including Political Theory (classical political philosophy and contemporary theoretical perspectives), Public Policy (the process by which governments make public decisions), Public Administration (the ways that government policies are implemented), and Public Law (the role of law and courts in the political process).

Empirical vs normative research

Political research has also tended to define a sharp distinction between 'empirical' and 'normative' research. **Empirical research** addresses events and political phenomena that we observe in the real world: questions about what *is*; **normative research** addresses questions about what should or *ought to be*. However, this distinction also appears to have decreasing relevance for many scholars. Empirical research is always shaped to some extent by normative thinking; and normative research embodies 'empirical' claims about the character of human and natural realities (Smith 2004: 86).

As John Gerring and Joshua Yesnowitz argue, empirical study of social phenomena 'is meaningless if it has no normative import; and it is misleading if its normative content is . . . [left] ambiguous', if we don't know how it matters (Gerring and Yesnowitz 2006: 104). Indeed, the justification for why a research question is interesting or relevant or meaningful is essentially a normative one. But at the same time normative arguments that propose or justify one value system over another will lack relevance if they make no attempt to relate to the facts of actual practice or public life. As Steve Buckler (2010: 156) points out, **normative theory** is concerned both about the world as it is and as we might think it ought to be. In sum, good social science is both empirically grounded 'and relevant to human concerns' (Gerring and Yesnowitz 2006: 133). Normative theorizing 'must deal in facts' and empirical work 'must deal in values' (2006: 108). We must be clear about the difference between normative and empirical questions and statements; but we must recognize that they are not independent of each other.

Recent discussions about the theory of deliberative democracy illustrate these points. Early research on the subject by scholars like Jürgen Habermas and John Rawls debated the normative justifications of deliberative democracy, interpretations, and necessary components of the theory, but failed to take account of the sheer complexity of contemporary societies (Elstub 2010: 291). However, recent research has tried to engage more seriously with the empirical realities of this social complexity (see Baber and Bartlett 2005; O'Flynn 2006; Parkinson 2006). As Chapter 3 will endeavour to show, **theory** and evidence inform

each other. Seeing either one as entirely divorced from the other generates either fantasy or mindless empiricism.

Questions of method continually arise in so-called normative research, and these are sometimes more analogous to those in empirical work than is always recognized. Normative theorists are concerned to convince others by means of drawing logical inferences and presenting the logical development of their ideas. They want to persuade others that the conclusions they reach are reasonable and plausible. An argument for any conclusion needs to be evaluated in light of the nature of its premises, the strength of the inferential links to the conclusion from these premises, and its possible criticisms or refutations.

Positivism vs interpretivism

However, studying 'political reality' is far from straightforward, and there is substantial disagreement about what methods can be used to analyse the social world. Debates over the respective strengths of different methods are often conflated with debates about what kind of *knowledge* can be gained about the social world. These are debates about epistemology, something we will be discussing in Chapter 2. **Epistemology** is concerned with understanding and explaining how we know what we know. It is important to understand something about the different positions scholars take with regard to this question because, *if we don't understand the different positions that exist with regard to what is knowable about the social world, we will be unaware of the implications of the choices we and others make when conducting research*. However, we argue that many of these different positions are also overstated when it comes to actually analysing the social world.

Many researchers have pointed to the tendency to exaggerate the differences between what scholars have tended to cast as 'two grand traditions in social science epistemology' (Pollins 2007: 93). These 'grand traditions' are represented by positivism and interpretivism, and the approaches that are based on them. **Positivism** maintains that scientific knowledge of the social world is limited to what can be observed; and that we can explain and predict social phenomena by discovering empirical regularities, formulating law-like generalizations, and establishing causal relationships. **Interpretivism** maintains that knowledge of the social world can be gained through interpreting the meanings which give people reasons for acting, and that we can, in this way, *understand* human behaviour, but we cannot *explain* or *predict* it on the basis of law-like generalizations and establishing the existence of causal relationships.

Positivism and interpretivism have different ontological and epistemological commitments: in other words, they have different views with regard to questions about the nature of the social world (**ontology**) and how we can have knowledge of it (epistemology). But in their actual research, those working in both traditions generally follow the same methodological conventions; and while researchers may be interested in different types of questions, 'practical investigation of these questions often leads them to similar methodological tasks' (Finnemore and Sikkink 2001: 395). Both are concerned to show the relations between premises and conclusions, and to indicate the nature of the relations between them. Both recognize that some standard of validation must be established for the sources of evidence used. By and large all researchers, whatever their methodological and philosophical

predispositions, share similar goals: to explain and understand, to engage with evidence and use it well, to distinguish between those claims about the world that are fanciful and those that are robust.

Our intention, therefore, is not to privilege any one way of doing research over any other. Instead, our view is that the method you use in conducting research will always depend on the answers to the following questions:

1. What research question are you trying to answer?
2. What evidence or data do you need to answer the question?
3. How are you going to analyse the data, and what are the practical steps needed to obtain and record them?

In our view, any research will be clearer, more accessible, more persuasive, and more likely to achieve its aims, if the researcher articulates a central question/puzzle/problem, provides a rationale for it, reviews the relevant literature, advances a hypothesis or argument, constructs a theoretical framework, defines concepts, variables, and relationships, and designs a 'test' of the hypothesis or argument. The distinctions drawn among and between qualitative and quantitative methods (to be discussed in the next sub-section), empirical and normative research, and positivism and interpretivism are important to understand for purposes of reflection on how to go about finding credible answers to important questions. However, while there is a tendency to draw a sharp a distinction between these types of research, we would emphasize that there is a common core of research practices that apply to all of them, that all research is judged by the same standard, and that no one type of research is better than another. It is important to the quality of your research to be aware of the issues raised in these debates. But these distinctions are ultimately sterile. Research is either systematic, self-aware, clear, and transparent, or it isn't. Irrespective of what type of research you pursue and what methods you use, your research will be judged according to some generally accepted standard of research practice.

There is one last distinction that we wish to address: the distinction between quantitative and qualitative political research.

Quantitative vs qualitative research

To reap the rewards of methodological diversity, while ensuring that research is more problem-driven and relevant to important political questions, requires that we develop a notion of some shared standard, one that can be applied to a variety of different questions and diverse research traditions. The search for a unified set of standards, or a common definition of what constitutes good research, has tended to become bogged down in a debate about whether and how it is possible to bridge the divide between quantitative and qualitative methods and approaches. **Quantitative research** tends to be based on the statistical analysis of carefully coded information for many cases or observations (in whatever way those observations are defined). **Qualitative research** tends to be based on the discursive analysis of more loosely coded information for just a few cases. Some even go as far as to say that quantitative and qualitative traditions are so ontologically distinct as to be

incommensurable. This position is based on a belief in a hard-and-fast connection between quantitative methods and the tenets of positivism, on the one hand, and qualitative methods and interpretivism, on the other. We think that this represents a false dichotomy. Different methodological positions are not tied to any epistemological or ontological position. Nor are they tied to any particular type of research question or research design.

Others suggest that some methods are inherently better than others. This position has frequently been attributed to the book *Designing Social Inquiry* (1994), by Gary King, Robert Keohane, and Sidney Verba, which began an important debate about methodology. Its aim, according to its authors, is to bridge the quantitative/qualitative divide (King, Keohane, and Verba 1994: 4). Its primary focus is causal inference in both qualitative and quantitative research and, more generally, scientific methods and elements of scientific research. The authors argue that 'All good research can be understood—indeed is best understood—to derive from the same underlying logic of inference. Both quantitative and qualitative research can be systematic and scientific', provided each submits to 'the rules of scientific inference—rules that are sometimes more clearly stated in the style of quantitative research' (1994: 4–5, 6).

Quantitative and qualitative research, the authors argue, are substantively the same, and differ only in the types of techniques they employ (1994: 4). Both are seeking scientific insights into social phenomena, and both require rigorous scientific method to ensure such results. But critics of *Designing Social Inquiry* argue that the authors develop a 'quantitative template for qualitative research'—one that presupposes the superiority of the former over the latter (Brady and Collier 2004: 3), and that this imposes positivist concepts like hypothesis testing on qualitative researchers.

We think some of these objections have been overstated. But we do reject the notion that some methods are better than others. Our view is that no method is inherently better than another, but only better or less suited to addressing the specific research question under investigation. Our concern, therefore, is not to prescribe *which* method or set of methods is to be preferred, but to explain *how* the various methods available to political researchers should be used. Consequently, with all the methods we discuss in this book, we focus on the sorts of questions that they may be used to address, and on their specific procedures, techniques, or strategies for collecting and analysing data or information necessary to demonstrate an argument. The link between theory and evidence is central to sound research, to how we actually go about collecting and analysing the information or evidence that we need to support (or just as importantly undermine) an argument. And while it is true that there is no such thing as a 'best' method, there is such a thing as convincing evidence and analysis, which is far more important. Evidence can be gathered, interpreted, and analysed in a variety of different ways. There are good ways and bad ways of doing this, irrespective of which method is used. No method is perfect. All have their strengths and weaknesses, and it is important to be aware of what they are.

The research process

There are a large number of texts available that address themselves to research methods, to specific aspects of the research process, and to debates relating to social science methodology: books on theories and methods, on research design, and on political analysis.

This book is also concerned with research design, analysis, and theory (though it does not devote chapters to specific theoretical approaches, such as rational choice or feminism). However, it is primarily designed as an introduction to the *research process*. The aim of this book is to introduce (1) some key issues in the philosophy of social science that bear on the choices researchers must make; (2) a step-by-step delineation of the process of asking and answering a research question—formulating a question, reviewing the relevant literature, advancing a hypothesis or argument, constructing a theoretical framework; defining concepts, variables, and relationships; and designing a 'test' of the hypothesis or argument; and (3) the array of methods used in political research and how scholars utilize these methods to answer questions about political phenomena.

Before elaborating on these aims, we wish to raise a point about the array of methods we introduce in Part 3 of the book. As we noted earlier, we are methodological pluralists. We are enthusiastic about the diversity of the field and about new methods of inquiry that are developing. But we do not cover every method and approach in this book. We doubt that it is possible to adequately survey, in a single volume, the vast array of methods and approaches available to political researchers; and we do not attempt to do so. Instead, we provide in-depth discussions of a range of widely used methods that we think can provide a base on which to build greater expertise and explore more specialized techniques; and we signal, where appropriate, how these can be further developed in more specific ways. Our Guide to Further Reading in each chapter directs readers both to specialist texts on the subject matter of the chapter and to more specialized adaptations of the approach discussed. We expect that students, with the help of the further reading for each chapter, will be able to take these approaches in a variety of further directions.

This book is concerned with the process and practice of political research: the principles and procedures that guide scholars as they conduct research, the beliefs and assumptions they hold, the kinds of questions they ask, and the variety of decisions that they must make. Its aim is to answer two questions:

1. How does one formulate research questions?
2. Once these questions are formulated, how does one design and carry out research in order to answer them?

To address these questions we focus on three broad components of the research process: (1) key issues in the philosophy of social science (the problems of what we can know and how we can know it); (2) the 'nuts and bolts' or the '*how to*' of research: *how to* find and formulate a research question; *how to* develop, through an engagement with the relevant literatures, an argument or answer that responds to your question; and (3) the specific methodological procedures and techniques utilized in carrying out a given research project. We will now provide an overview of each of these components.

Part 1: Philosophy of Social Science: Knowledge and Knowing in Social Science Research

Methodology refers to the conduct of inquiry. Among other things, this involves reflection upon the system of values and principles that guide social analysis. Questions of ontology and epistemology—questions about the complexities and ambiguities of knowing and

gaining knowledge of the social world—are a core part of this reflection. These questions are the focus of key debates in the philosophy of social science and, in our field, of ongoing debates about scientific practice, forms of knowledge, and the world of politics.

Why, you might ask, should these philosophical issues and controversies concern political researchers? To start with, it is important to understand that all social scientific theorizing adopts a position with respect to the issues that these controversies involve. If you are unaware of them, you will not fully understand the implicit assumptions and implications, either of what others are arguing, or of what you yourself are arguing. You may make statements in which you are unknowingly committing yourself to assumptions or principles with which you disagree and which are inconsistent with the position or argument you wish to advance. We have said that research methods build on and develop your own natural abilities to think and solve problems. In the same way, an understanding of philosophical positions concerning knowledge helps to make explicit and to develop the philosophical assumptions about, and approach to, knowledge that you already employ in your everyday life.

Furthermore, if we see the purpose of scholarly research as the generation of valid and reliable knowledge, and view truth as a central characteristic of valid knowledge, then we need to adopt and defend assumptions about the nature of truth and how to discover it. In other words, we need to have a sound understanding of how to establish that something is true. Philosophical presuppositions about 'reality' are thus intrinsically linked to the substantive issues that are central to inquiry in our field. Researchers cannot contribute to knowledge about something unless they adopt and defend a view of what is knowable about the social world and what sorts of things different methods enable us to know.

Philosophy of social science debates have implications for all areas of research; and, whether or not researchers follow ongoing debates in the philosophy of social science, their views tend, either implicitly or explicitly, to reflect one or another of the different answers and positions that these debates have generated. All of us have lodged in our minds a 'world view' which shapes our opinions about what goes on around us. This is constituted, most likely in large part, by 'an accumulation of the ideas and prejudices of others' (Kahane and Cavender 2006: 19). If the unexamined life is not worth living, as Socrates is said to have claimed, then it perhaps follows that unexamined beliefs are probably not worth holding. Learning how to systematically investigate ideas we hold about the world enables us to analytically engage with political affairs, rather than to passively consume the output of politicians, political analysts, and the news media. Reality is constantly being defined for us. The ability to identify the underlying structure of assumptions which shapes a given account of reality, whether presented by scholars, politicians, or journalists, allows us to become more active analysts of contemporary politics.

In sum: understanding the terms of the major debates in the philosophy of social science, and sensitivity to their implications, is an important part of producing good research. What you presume is knowable about the social world will bear on the strategic choices you will need to make all through the process of research.

Part 1 of the book is devoted to two key controversies in the philosophy of social science: (1) how we can know and develop knowledge about the social world, and (2) whether it is possible to separate facts and values in social science research.

The first controversy involves questions about what sort of knowledge we can gain about the social world. Is it the same sort of knowledge that scientists are able to obtain about the natural world? Or are the forms of knowledge concerning the social world and the natural

world necessarily different? In Chapter 2 we address three different approaches to answering these questions and their implications for conducting political research. Two of these approaches have already been briefly discussed in this chapter.

The first is 'positivism'. Positivism is usually defined by the following three tenets: (1) scientific methods (i.e. the testing of hypotheses derived from pre-existing theories) may be applied to the study of social life; (2) knowledge is only generated through observation (empiricism); and (3) facts and values are distinct, thus making objective inquiry possible (Snape and Spencer 2006). The second approach, 'interpretivism', maintains that the social world is fundamentally different from the world of natural phenomena, and that it does not exist independently of our interpretation of it. The task of social science, then, is fundamentally different from that of natural science, because the objects of the social sciences are different from those found in the natural world. The third approach discussed in Chapter 2, 'scientific realism', is similar to positivism in some ways because it accepts some positivist assumptions. However, while positivists maintain that reality consists of only that which we can directly observe, scientific realists maintain that reality consists of both observable and *unobservable* elements. Scientific research is therefore based, not only on evidence that is visible in the world, but on unobservable facts, objects, or properties that exist independently of our direct knowledge of them. Consequently, the goal of scientific inquiry is to describe and explain both observable and unobservable aspects of the world.

Just above, we identified the three tenets of positivism. The third tenet maintains that it is possible to define a distinction between facts and values, and for us to acquire value-neutral, objective knowledge about social phenomena. However, many in our field dispute this claim. They argue that knowledge produced by social-scientific research is not value-neutral, but is shaped by a variety of factors, including existing scientific theory, politics, and power relations, cultural beliefs and meanings, and the researcher's own motivations and values. What do you think? Can knowledge produced through the study of the social world be 'objective'? We consider this question in Chapter 3.

The position you ultimately arrive at with regard to the issues raised in each of these debates have important implications for how you pursue research and develop explanations of political phenomena.

Part 2: How to do Research: An Overview

A second component of the research process involves the basic nuts and bolts of the research enterprise. By this we mean the steps involved in developing a plan for pursuing research on a topic. This involves developing a researchable question, locating applicable theory and literature, formulating testable hypotheses, and clarifying concepts and developing empirical indicators. The chapters in Part 2 of the book introduce basic elements of research: a research question, what requirements it must meet, where to find and how to formulate one (Chapter 4); how to construct an argument, hypothesis, or theory that answers it (Chapter 5); how to design research to test, investigate, or demonstrate it (Chapter 6); and the information or data we collect for our analysis (Chapter 7).

Most political research originates from some general question or problem that arises, either from the events, issues, or processes we observe in the world around us, or from the

theories and frameworks that our field has developed in order to understand them. But the research *process* only begins with the conversion of this general question or problem into a well-formulated, clearly focused research question. *Step 1* of the research process involves finding and formulating a researchable question, and locating applicable theory and literature (Chapter 4). We discuss why research should be guided by well-formulated research questions, the role of the 'literature review' in providing both the inspiration and the rationale for them, and the different types of questions scholars ask. We discuss the basic requirements of a research question and how to go about meeting them. In our view, a research question is one that (1) has significance for a topic or issue relating to the subject matter of our field, (2) is researchable (it can be answered through conducting research), and (3) has not yet been answered definitively.

Irrespective of the particular research one chooses to undertake, a carefully formulated research question will ensure that the research has a clear purpose with clear goals. A research question makes the research task specific so that, rather than trying to gather all the information you can find on a topic, you direct your attention to just the information that addresses the question and helps you to explore the argument you develop to answer it. A well-formulated research question helps you make decisions about what information should be included or excluded, what data you need to collect, what to observe, or what to ask your interviewees.

To formulate a research question you need to be clear about what you want to know. We identify different types of questions, what each type commits you to doing, and the logical fallacies that, if they find their way into the formulation of a question, can lead to false conclusions. The third requirement of a research question requires that you show that others who have addressed it have failed in some way to provide a definitive answer to it. The literature review performs this function. It identifies what (range of) answers are found in the existing literature relating to the question and it develops an argument about what needs to be done to provide a better answer to the question than those that currently exist.

Once you have formulated a research question, you will be ready to move on to *Step 2* of the research process: how to answer it. We discuss the basic components of an answer to a research question, what requirements it must meet, and where to find and how to formulate one (Chapter 5). To get you started we offer a broad template. This is meant to serve as a starting point—a set of considerations to be revised to fit with the question and aims that animate a particular project. But though the template shows a research process that is structured in an idealized linear fashion, we recognize that the process as it actually unfolds is often not linear at all. The real process is often circuitous. Most researchers move back and forth between theory and evidence—between theorizing something that is the focus of their research, and mucking about in the 'dust of detail' (learning more about the specific facts of the case or question or issue, or the observations of other scholars that we treat as facts). As Philip Shively observes, one of the better-kept secrets' in our field is that good researchers usually do not

> 'frame hypotheses' in any formal sense before they start to work, though they may have some operational hunches about what they expect to find . . . They play with data, immerse themselves in what other people have written, argue with colleagues, and think.
>
> (Shively 1989: 25)

In sum, we describe the research process as consisting of a set of components or series of steps but, in practice, the process of research does not unfold in the sort of linear fashion that this suggests. We re-think our views as a result of learning from the research process itself in ways that can feed back to our previous choices and lead us to revise them. So, what the template we present really shows is not the *process* of research, but its ultimate *presentation*.

We organize our discussion of the various considerations and tasks involved in developing an answer to a research question around three basic requirements. The *first* requirement is that the answer be appropriate to the type of question that is being asked. Different types of questions demand different types of answers. For instance, *descriptive* questions will require that you describe the characteristics of something, or model how it works or behaves. *Explanatory* questions will require that you explain what factors or conditions are causally connected to a known outcome. *Normative* questions may require that you adjudicate among different understandings of how something should be, or what should or ought to be done, by considering the arguments of others, and submitting well-reasoned arguments for your own.

The *second* requirement of an answer to a research question is that it makes a contribution to knowledge. Social science research is expected to address a question whose answer will contribute to collective knowledge in a particular field of study; so in developing an answer to a research question one must ask oneself: Why should we care about this answer or argument? In other words, your answer must matter. It is not enough to say that the question has not been asked before. After all, one very good reason why a question has not been asked could be because no one cares about the answer. So you must always make a case for why the question is relevant and important.

The third requirement is that an answer must be clearly and fully specified with regard to the factors or variables you think must be taken into consideration in order to answer your question, and how you think these factors or variables are related to each other. For all types of research, we think it is useful to formulate a 'working hypothesis'—an operational hunch about what you expect to find. Initially, what argument motivates the research? What findings might be expected? Because a hypothesis articulates *in advance* the contours and logic of the investigation, it helps to guide research. The term 'hypothesis' is often treated as applicable only to quantitative research and to a specific prediction about the nature and direction of the relationship between two variables. But we use the term 'hypothesis' to mean 'a hunch, assumption, suspicion, assertion, or idea about a phenomenon, relationship, or situation', with which research begins and which becomes the basis of inquiry (Kumar 2005: 74). The key thing is that the hypothesis should be empirically or logically verifiable (or not). That is, it should be falsifiable with evidence. A hypothesis which is not falsifiable is really just a tautology, and so is not going to tell us anything interesting about the political world.

If we are interested in the question we ask and the hypothesis we test, then we should also be interested in the answer we get; and that means that an answer which discredits our hypothesis is just as valid and interesting as one that confirms it. For example, the finding that the level of ethnic diversity in a country *does not* influence the level of democracy in the country has important implications for our understanding of what impedes democracy (and what does not).

Hypotheses can either be tested with evidence (**confirmatory research**), or operate as a guide to a process of discovery (**exploratory research**). Exploratory research begins with

a question and perhaps a basic proposition, probes its plausibility against various types of data, and eventually generates a more concrete hypothesis, which can be more widely and rigorously tested.

Once you have a hunch or argument about the answer to your research question, you then need to develop a strategy for providing a convincing 'test' or demonstration of it. This is *Step 3* of the research process: how to demonstrate the validity of your answer (Chapter 6). The plan you develop to do this is what we call a **research design**. The research design sets out a plan for research, including what observations to make and how to make them; it is informed by and fulfils a logical structure of inquiry. It specifies the sort of test or evidence that will convincingly confirm or disconfirm a hypothesis; the observations you need in order to demonstrate the relationships stated by your hypothesis, and how you will go about making them; and the data relevant to demonstrating these relationships, and how and where you will collect them. This chapter previews a variety of research designs that are used by scholars in our field. The type of research design you use and the kind of information you collect, the sources of data and the data-collection procedures you choose, should be based on what will provide a convincing test, demonstration, or investigation of your argument or hypothesis.

Among the considerations that researchers need to take into account when choosing a type of research design are the ethical factors that may be involved in the sources and data-collection procedures you intend to use. We present a series of short discussions about

BOX 1.1 Key Terms

Empirical Research	Research based on first-hand gathering of data through, for instance, interviews or surveys.
Quantitative Research	Empirical research in which the researcher explores relationships using numeric data.
Qualitative Research	Empirical research in which the researcher explores relationships using textual, rather than quantitative data.
Normative Research	Research that addresses questions about what *should* or *ought to be*; that seeks, for instance, to propose or justify one value system over another.
Research Design	A procedural plan which the researcher develops in order to answer research questions validly, objectively, accurately and economically.
Confirmatory Research	Research which tests a hypothesis with evidence.
Exploratory Research	Research which operates as a process of discovery, in which the plausibility of a hypothesis is probed against various types of data so as to eventually generate a more concrete hypothesis which can be rigorously tested.
Positivism	An epistemological stance that maintains that all phenomena, including social, can be analysed using scientific methods. Everything can be measured and, with enough knowledge, the causes and effects of all phenomena could be uncovered.
Intepretivism	A standpoint that rejects the assertion that human behaviour can be codified in laws by identifying underlying regularities, and that society can be studied from a detached, objective, and impartial viewpoint by the researcher.

ethics in research across a number of chapters in Part 3. We begin our series of discussions in this chapter with a consideration of the ethical principles to which all research should conform.

We discuss the principles of what makes good data in Chapter 7. If we want to develop our understanding about the political world or investigate different ideas or explanations about how things work, we need to think about how our ideas relate to what happens in practice. This means engaging with the empirical world—and to do this we need to collect information. Our record of the information that we have collected is data. We then use this data as evidence to support our arguments. Data thus constitutes the information base for our analysis. Without data—or evidence—argument is little more than conjecture. In Chapter 7 we discuss the main types of data that are used in politics and international relations, and the main criteria by which we can judge whether our data is good or not. Whatever type of data we use in our analysis, there are a number of issues that we must confront. First, as far as possible we need to ensure that our data is valid: that is, the data does actually measure what we claim it does. And, second, we need to ensure that our data is reliable: that is, we have recorded the data (or information) accurately.

Before moving on to a discussion of Part 3 of the book, take a look at Box 1.1, above, which reviews terms we have introduced in this chapter thus far. These will be the focus of considerable discussion throughout the book.

Part 3: How to do Research in Practice

The third part of the book is devoted to examining the principles of different methods in political research, and how to use them in practice. These methods enable us to put a research design into practice and to carry out the research and analysis needed to answer a research question. Methodological principles are concerned with *how we obtain knowledge* about the political world—with the specific procedures or techniques that we use to tell us something about the world when we carry out research. Research methods are in a sense the tools of analysis. And, as with any tool, it is important to know how to use it in order to complete a task.

In Part 3 of the book we focus on different strategies for collecting and analysing evidence in order to test or develop specific hypotheses. In particular, we focus on the ways in which data is collected and the ways in which it can then be analysed. It pays to think of these separately, because these are two distinct and crucial aspects of the research process. Failure to devote sufficient care and attention to either will lead to research results which are untrustworthy and unreliable, meaning that we may reject hypotheses that are actually true or accept hypotheses that are actually false. Obviously, we want to avoid the possibility of either of these outcomes occurring as far as possible. If we reject or accept a hypothesis, we want it to be because we have collected convincing data and analysed the data properly.

Different methods of **data collection** include collecting information through the use of experiments (Chapter 8), comparative research (Chapter 9), historical research (Chapter 10), surveys (Chapter 11), interviews or focus groups (Chapter 12), participant observation (Chapter 13), or collecting archival data or documentary records such as speeches, policy documents, or media reports (Chapter 14). The type of evidence that we

use is important. But more important is how we collect and code the evidence in preparation for analysis. This relates to the issues of validity and reliability. We want the data we have collected to be reliable and robust. It is what separates good research from reliance on rumour, hearsay, and anecdotes. Data collection needs to be systematic, not haphazard. We must have guiding principles to ensure that the data we collect is of good quality.

Having obtained this data or information, we can then think about how to go about analysing it. It is common to distinguish between **quantitative** and **qualitative** approaches as methods of analysing data. Quantitative analysis tends to rely on statistical techniques, which we discuss in detail in Chapters 15, 16, and 17. But there is a wide variety of more qualitative analysis techniques that we can consider. It is worth remembering that this distinction between qualitative and quantitative analysis primarily refers to different methods of analysis rather than to different methods of data collection. And although there is some overlap between how we collect evidence and how we analyse it, there is also a lot of diversity, and similar types of data can be analysed in different ways. In fact data from all the types of data-collection methods that we discuss can be analysed using either quantitative or qualitative techniques. When we talk about 'quantitative data', we are only talking about evidence that has been coded in preparation for quantitative analysis. This involves coding and categorizing the data with numerical values. The actual form of the evidence can be almost anything. It is therefore important to be familiar with both qualitative and quantitative approaches, since failure to understand or engage with one approach is likely to cut you off from a lot of relevant research that is carried out on your topic of interest.

Much is often made about the relative strengths and weaknesses of different methods, in particular between quantitative and qualitative approaches. It is frequently asserted that quantitative research may be good for making generalizations, but it is a blunt instrument for investigating hard-to-define concepts such as power, globalization, and democracy, or difficult-to-observe phenomena like money laundering, criminal or anti-social behaviour, corruption, and terrorism. By contrast, one of the key strengths of qualitative research is often thought to be its ability to investigate these hard-to-define concepts and hard-to-reach populations. However, whereas qualitative research might be able to go into a lot of detail, there is sometimes a nagging suspicion that its findings might not have wider relevance to contexts outside the immediate vicinity of where the research was conducted. While there might be an element of truth in this, these strengths and weaknesses should not be overstated or viewed as inherent in the different methodological approaches.

The strengths and weaknesses of qualitative and quantitative approaches are often presented in terms of a trade-off between description and detail (validity of measurement), and explanation and generalization (validity of inference). We reject this idea. The ability to make generalizations is not just the preserve of quantitative approaches, but about something more fundamental to the research process. It is about being able to rule out and control for theoretically plausible alternatives. All methods of analysis should take this issue seriously, otherwise we may end up with spurious findings. The only way to address this is by considering theoretically important variables; and to do this requires that we consider enough cases to make it manageable. For example, the concern that the findings from a *small **N*** (where *N* refers to *number*, as in number of cases or countries, etc.) study may not be applied more generally are to do with concern over omitted variable bias, which can cause spurious relationships. This we discuss in detail in Chapter 9.

To be able to estimate the effect of many variables you need lots of cases to see if your argument still holds up when you consider different factors across different contexts. Quantitative methods provide an efficient way of doing this. But qualitative methods can do it as well. And indeed they often do, as we discuss throughout the book. But it is generally a slower and less systematic process which takes a great deal of time. Studies are repeated, findings are applied to new contexts, and, as this is done, theories or hypotheses are indirectly tested on a wide range of cases. But there is nothing to stop it being done quicker and more systematically; it would just involve more money. The difference is not one of method but of time and costs. There is no methodological reason to stop you repeating qualitative studies; you can do twenty in-depth interviews (see Chapter 12); but why stop at twenty? If time and resources are no obstacle, then it is possible to do 2,000 interviews. And why not collect multiple ethnographies? Recent research is attempting to do just this, through collaborative comparative ethnographies that try and develop a firmer basis for wider inferences (see Gillespie, Gow, and Hoskins 2007; Gillespie and O'Loughlin 2009).

It is often said that qualitative methods can provide more valid measures of political phenomena. They are better able to measure and study difficult-to-define political phenomena, whereas quantitative methods are too blunt and reduce complicated concepts down to numbers, which can never capture the full meaning of what is being investigated. But this is not the case either. There may be an issue to do with the extent to which extant survey data can be used to answer new questions, but the answer to this is to design new surveys or collect new data and devote additional resources to the problem. The difference is one not of method, but of resources. It is possible to measure anything in a reliable and valid way with quantitative methods of data collection, but to do so if the object of investigation is difficult to observe or measure can be very costly. Studies using qualitative methods are often relatively cheap.

Our position is that, irrespective of whether a quantitative or qualitative approach is employed, the questions, decisions, concerns, and procedures with which researchers have to deal in designing research are similar. The main considerations are always to do with sampling and measurement. Whether we are doing quantitative research or qualitative research, we want the answers we arrive at to be meaningful and say something about the world. We don't want them to just reflect the way in which we have chosen to measure political phenomena or selected particular cases to analyse.

In discussing these different forms of data collection and **data analysis**, we draw on some of the most prominent approaches in political research. We do not discuss every method used in political research, but a selection of prominently used ones that can serve to introduce themes and research protocols generalizable to other methods. Each chapter will provide examples of how published research actually employs these methods or approaches. A number of methods will require some consideration of ethical issues, and we will flag this up where relevant.

We start this section with an introduction to the principles of experimental research in Chapter 8. The experimental approach is widely considered to be the most 'scientific' research design. Through the use of **control groups** and **experimental groups** the researcher is able to control what stimuli—or interventions—different subjects are exposed to, and then examine what impact this exposure has on the political outcome variable of interest. Controlled experiments of this type are very useful for testing causal hypotheses. Broadly

speaking, there are three main experimental designs. There are **laboratory experiments** (where subjects are taken to a common location), **field experiments** (which take place in real-world settings), and **natural experiments** (which in a sense occur naturally, in so far as the researcher is not active in the data-gathering process). Despite the scientific potential of experimental research, the approach is not widely used in the study of politics. However, this is beginning to change. Experiments are now one of the fastest growing fields of political inquiry. We discuss some of the obstacles that have been traditionally associated with doing experiments in political research, and discuss the potential for developing and expanding their application.

In Chapter 9 we focus on comparative research. Comparative research represents one of the largest fields of political inquiry, and to a certain extent is used by all investigators who engage in empirical research. The comparative method (or approach or design, as it is sometimes termed) actually involves a number of different methods and can be used in conjunction with any method of data collection. The logic of comparison is based on how many countries (or cases) are compared, and how the cases for analysis are selected. Both aspects of case selection are very important, as the cases you look at can affect the answers you get to any particular research question (as Barbara Geddes 1990 has pointed out). Broadly speaking, there are three main approaches. There are **large-N** studies (involving the analysis of many cases), **small-N** studies (involving the analysis of a small number of cases, typically two, three, or four, but with no real upper limit), and **single-N** studies (otherwise known as **case studies**). Virtually all political research falls into one of these three subtypes. One of the key strengths of comparative research, particularly when it involves the analysis of several or more countries, is that it provides a bridge between looking at domestic factors (which take place within countries) and international factors (which take place between countries). Comparison helps us to broaden our intellectual horizons, and we can use comparison to see if what we think is a self-evident truth in one context also works in the same way in a different context.

Chapter 10 focuses on historical research. Political research and history have always come together in key areas of inquiry in our field: in the study of the origin and development of states and political institutions, the requisites of democracy and of development, and the dynamics of social classes and of revolution; as well as in international security, international political economy, and foreign policy analysis. In recent years, however, researchers in our field have been increasingly drawn to historical approaches because they enable us to incorporate a stronger temporal dimension into our analyses. Analysis becomes *temporal* by situating events in their historical and cultural contexts, taking into account related series of events that unfold over time, and thinking explicitly about timing and sequence. There are two forms of temporality. Temporality can be conceived of in terms of (1) a broad historical period (time as context), and also of (2) the sequential active unfolding of social action and events. Each of these forms underpins a set of methods. The first form of temporality, time as context, is represented by various methods of comparison which were discussed in Chapter 9, but which are further elaborated in Chapter 10. The second form of temporality is time as the sequential unfolding of social action and events in and through time and places. This type is represented by historical institutionalist analysis and its associated methodologies, and **event structure analysis** (ESA). Historical institutionalism is concerned with bringing questions of timing and temporality to the centre of the analysis

of institutions. ESA is an analytic procedure designed to help researchers in 'unpacking' an event 'and analytically reconstituting it as a causal interpretation of what happened and why it happened as it did' (Griffin 2007: 5).

The following chapters explore various methods of data collection and analysis. Among the most widely used forms of data collection in political research are **surveys**, which we discuss in detail in Chapter 11. One of the great strengths of survey research is that it helps us to make general claims about what different sections of society or different sub-groups of the population actually think and do. It thus gives voice to people who might not otherwise be heard. But a major weakness is that surveys can, and frequently do, misrepresent what people think and do, and thus create misleading information. The extent to which surveys misrepresent the 'real' or 'true' attitudes and behaviour of the people they seek to study can be thought of as error. The purpose of a good survey is to try and minimize this error. There are two important sources of error that we consider. The first is to do with **measurement error**, and refers to the ways in which surveys use questions to try and measure different social and political phenomena, such as political attitudes, opinions, and behaviour. The second is to do with **sampling error**, and refers to the ways in which respondents are chosen or selected to complete the survey and the implications this has for the representativeness of the sample. These principles of sound survey design, to do with sampling and measurement, are relevant for all forms of data collection.

In Chapter 12 we focus on data collection using **interviews** and **focus groups** and explore issues concerning how these data can be analysed. Interviews are in a sense the qualitative cousin of surveys. Many of the principles are much the same for the two methods of data collection, but whereas surveys are typically concerned with generating large samples so that they can make valid inferences about a given population, interviews are more frequently used to ascertain more specialized knowledge, either about what so-called experts or elites think, or to explore the meanings that people attach to different concepts. It can be a very useful method to complement survey research, and indeed virtually all survey research draws upon semi-structured interview techniques at the design stage to pilot new questions. One of the great strengths of interviews and focus groups is that they can help a researcher understand people's perceptions, feelings, opinions, experiences, understandings, values, beliefs, attitudes, emotions, behaviour, formal and informal roles, and relationships. Interviewing individuals, either face to face or over the telephone, or through mailed questionnaires, helps researchers to learn about how people feel. The focus group is a good technique for exploring *why* people hold certain beliefs or feel the way they do. Exchanges among participants can lead to far more probing and reflection than is possible in individual interviews or questionnaires, and may provide more robust and revealing responses to the issues which are the subject of the focus group.

In Chapter 13 we consider participant observation, the most intensive form of data collection of all. The distinctive feature of participant observation, and one of the great strengths of the approach, is that data collection is carried out in real time. This means that the researcher has a direct, first-hand opportunity to observe what people actually do, what they actually say to each other, and how they actually interact with different institutions or political processes, rather than just relying on what people say they do. Participant observation (and ethnography more generally) therefore has a number of characteristics that overlap with other methods we consider (such as surveys, focus groups, and interviews) and a

number of characteristics that are distinctive, particularly with respect to the role of observation. Whereas surveys are based on the ancient art of asking questions to find out what people think, say, and do, participant observation is based on something rather different. It recognizes that what people say they do, and what they actually do, can be and frequently are quite different. Accordingly, to get a 'true' sense of what people think and say and do, it is not enough to merely ask people questions and record their answers; it is also necessary to observe what people do in practice.

In Chapter 14 we consider a different type of data collection. Whereas surveys, interviews, and participant observation are all to do with collecting information about what people think or say or do, either by asking them questions or by observing what they do, or some combination of the two, textual analysis is based on the analysis of archival data or documentary records, such as speeches, policy documents, or media reports, to do with what people or institutions or organizations have actually done (or produced). These documents provide a rich source of information about the ways in which politics is practised. We discuss some of the different ways in which these sources of evidence are analysed, focusing on historical analysis, discourse analysis, and content analysis. Unlike asking people questions (e.g. through surveys or in interviews), using texts to collect data has the advantage of being non-intrusive. Researchers do not face the problem of influencing their data source through the questions they ask. And they can study past policy positions as they were recorded at the time. Once recorded, texts do not change.

Quantitative analysis is now one of the most widely used techniques of analysis in political research. Whether you love it or hate it, it is hard to avoid. And although many students are apprehensive about quantitative analysis, it is an important skill to acquire, which will not only stand you in good stead for conducting research (both in terms of what you can read and what you can do), but will also provide you with a transferable skill. In Chapters 15, 16, and 17 we provide a step-by-step guide to quantitative analysis, which will equip you with the skills to be able to interpret what others have done, and carry out quantitative analysis for yourselves.

Although it is easy to overstate the differences between these different methods, it is also important to bear in mind their similarities. All methods of data collection and analysis can (and frequently do) investigate similar political phenomena. There is no sense in which the investigation of certain topics or issues determines your method of inquiry. For example, if you are interested in what people think about politics, you can examine this through the use of surveys, interviews, and focus groups, and ask questions about what people think and say they have done. You can also explore this through ethnography and participant observation and record what they actually do and say. Rather than thinking about how to choose between different forms of data collection and analysis, the more pertinent issue is often to think about how to combine different methods.

Researchers frequently combine different methods. Using different data sources or research methods generates evidence that can increase the overall validity of their research findings. This, as we explain in Chapter 6, is what researchers call '**triangulation**'. An example of triangulation would be a study that incorporated surveys, interviews, and statistical analysis.

Combining different methods in one study is more generally described as a '**mixed-methods approach**'. Combining different methods in one study can compensate for the

limitations of each method. By generating multiple perspectives on a phenomenon, it broadens the scope of investigation. This enables researchers to investigate a research question more comprehensively, and it can improve the way we think about and study complex political phenomena.

A mixed methods approach frequently used in political research is one which combines formal models, quantitative, and qualitative methods. Lisa Martin (1994) uses this 'tripartite approach' (Laitin 2002) in her study of international economic sanctions. In her book, *Coercive Cooperation: Explaining Multilateral Economic Sanctions*, she explores the conditions necessary for the successful imposition of sanctions. Success in this depends on cooperation; so what, she asks, makes cooperation to impose sanctions more likely among the trading partners of the target state? To investigate this question, Martin uses game theory, statistical analysis, and case studies. Game theoretical models yields two hypotheses about the decision process of countries considering whether to cooperate to impose economic sanctions on a third country: that cooperation to impose sanctions is more likely (1) against a weak, poor target country and (2) where international institutions are involved. To investigate the hypothesized correlation between the involvement of international institutions and cooperation she uses a statistical analysis of ninety cases of international cooperation on economic sanctions. She then uses case studies to help evaluate both the hypotheses generated by the game theoretical models and the correlations offered by the statistical analysis. In sum, Martin first offers an analytical framework using formal methods, then establishes statistical associations between variables, and finally investigates case studies that can either confirm or falsify the proposed hypotheses.

We present a more extended consideration of mixed methods in Chapter 6, along with a 'Research in Action' box: a discussion of a piece of published work which employs mixed methods and how the author uses them to provide a robust answer to a research question. Many of the chapters in Part 3 of the book include this 'Research in Action' feature. Please note that what we mean by 'Research in Action' is not to be confused with something called 'Action Research' in Politics and International Relations. 'Action Research' is concerned, not with methods or approaches, but with how the researcher themself might become more publicly engaged with the events, processes, and groups which make up everyday politics (see for a discussion of this, Putnam 2003; Macedo 2005; Gibbons 2006; ISR 2011). Our 'Research in Action' boxes provide a discussion of a piece of published research which illustrates a particular method or approach.

There is no rule about which method should be used for which research design. From Table 1.1 we can see that there is considerable overlap between different research designs, methods of data collection, and methods of data analysis. For example, longitudinal studies may use qualitative historical analysis or quantitative statistical analysis. There is also considerable overlap between different methods of analysis and different methods of data collection. Although it is common to associate quantitative analysis with survey research, there is nothing to prevent this type of analysis being carried out on data collected using other methods. Indeed, a great deal of quantitative analysis has been done on interview data based on expert or **elite interviews** (such as Polity IV, Freedom House, or Transparency International data), and on documentary records and public records and media reports (see Chapter 12). Even focus groups can be analysed quantitatively if there are enough of them. This approach is frequently adopted in deliberative polling studies (see Luskin, Fishkin,

Table 1.1 Research design, data collection, and data analysis in political research

	Research design					Method of analysis	
	Experimental	Comparative	Cross-sectional	Historical	Longitudinal	Quantitative	Qualitative
Method of data collection							
Surveys	✓	✓	✓		✓	✓	✓
Interviews	✓	✓	✓		✓	✓	✓
Focus groups	✓	✓	✓		✓	✓	✓
Participant observation	✓	✓	✓		✓	✗	✓
Texts	✓	✓	✓	✓	✓	✓	✓
Method of data analysis							
Quantitative	✓	✓	✓	✓	✓	–	–
Qualitative	✓	✓	✓	✓	✓	–	–

and Jowell 2002), which can be regarded as a type of experimental focus group. Indeed it is only really participant observation that does not appear to lend itself to quantitative analysis, though that is not to say that participant observation doesn't incorporate quantitative analysis or that it cannot be combined with quantitative analysis.

This all goes to show that in some respects the opposition between qualitative and quantitative approaches have been somewhat overstated. These should be seen, not as competing approaches, but as complementary approaches to political research. Both can be used for similar research designs, and to analyse similar types of data. It is not the case that there are some types of data that you can only analyse qualitatively or some types of data that you can only analyse quantitatively. There is, in fact, considerable overlap. So if you are interested in conducting a longitudinal study to investigate changing patterns in the tone of political coverage in newspapers, why not employ both qualitative and quantitative research? If different methods of data collection can be employed for similar research designs, and different methods of data analysis can be employed for similar forms of data, then the question is not why you should choose one method of analysis or collection over another, but why you should not choose all over one.

Unfortunately, the answer is usually pragmatic rather than intellectual, and has more to do with constraints of time, money, and expertise. For example, participant observation can be very time-consuming and surveys can be very expensive. Moreover, once researchers have acquired one methodological skill-set, they are often reluctant to learn another. These constraints are not irrelevant, but they should be openly acknowledged rather than obscured by pseudo-methodological, epistemological, or even ontological arguments. We hope this book will enable you to be comfortable using a wide variety of different methodological approaches and skills.

Research methods are in a sense tools of analysis. It is common to think about different methods being a bit like different tools. Which tool you choose will depend upon the job you want to do: whereas one tool might be appropriate for one type of task, it might be

inappropriate for another. So a saw is good for cutting things, but less helpful if you want to fix something together. Although there is some truth in this metaphor, we think the idea of choosing different methods according to the job you want to do leads to a very narrow view of what constitutes research. We prefer to look at the bigger picture. No one wants to carry around a saw and look for things to cut in half all the time. Rather, a skilled carpenter may want to build a table and to do that they will need to cut things and fix things and sand things and varnish things. They will need to use a wide variety of different tools to complete their task. Becoming a skilled researcher is a bit like learning a craft. And it involves learning many skills and being able to combine them together to do a thorough job.

Of course, for the lazy or time-pressed there is always flat-pack furniture. If we buy a flat-pack table from Ikea, we can assemble it all using a single tool—just an Allen key or a screwdriver is needed. But the quality of the table is not going to be in the same league as one hand-built by a master craftsperson, and it may well collapse under a slight amount of pressure. If we don't want our research findings to collapse in a similar way, we should aim to set our sights a little higher than trying to investigate complicated political phenomena through the prism of a single methodological perspective.

Conclusions

One of the great strengths of political research is its diversity. The study of politics encompasses a wide variety of different ontological, epistemological, and methodological positions. And while it might be disconcerting or confusing that there are no concrete shared principles that govern the study of politics, this variety is a great source of vitality in the discipline. Different perspectives act to continually challenge what we think we know. Constant questioning means that we can never take things for granted. We always need to be conscious of whether or not we can defend ourselves from criticism, since our findings will be scrutinized on many different fronts.

One of the defining characteristics of politics as a field of study is the wide variety of approaches it incorporates. This can be both a strength and a weakness for the discipline. It is a strength when this diversity is embraced, and when researchers adopt and integrate the different approaches and engage with research from across the methodological spectrum. It is a weakness when this diversity fragments the field, and when researchers from different methodological traditions retreat into their own enclaves and do not engage with what other people are doing in the discipline. Our view is that, to appreciate the diversity and pluralism within political research, it is useful to be familiar and conversant with the whole array of methods and approaches available to us. Our hope is that this book will help you in this task.

References

Baber, W. and R. Bartlett (2005), *Deliberative Environmental Politics: Democracy and Ecological Rationality* (Cambridge, MA: MIT Press).

Brady, H. and D. Collier (eds) (2004), *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Lanham, MD: Rowman & Littlefield).

Buckler, S. (2010), 'Normative Theory', in D. Marsh and G. Stoker (eds), *Theory and Methods in Political Science*, 3rd edition (Basingstoke: Palgrave Macmillan), 156–80.

Elstub, S. (2010), 'The Third Generation of Deliberative Democracy', *Political Studies Review* 8(3): 291–307.

Finnemore, M. and K. Sikkink (2001), 'Taking Stock: The Constructivist Research Program in International Relations and Comparative Politics', *Annual Review of Political Science* 4: 391–416.
Geddes, B. (1990), 'How the Cases you Choose Affect the Answers you Get: Selection Bias in Comparative Politics', *Political Analysis* 2(1): 131–50.
Gerring, J. and J. Yesnowitz (2006), 'A Normative Turn in Political Science?', *Polity* 38(1): 101–33.
Gibbons, Michael T. (2006), 'Hermeneutics, Political Inquiry, and Practical Reason: An Evolving Challenge to Political Science', *The American Political Science Review* 100(4) (November): 563–71.
Gillespie, M. and B. O'Loughlin (2009), 'News Media, Threats and Insecurities: An Ethnographic Approach', *Cambridge Review of International Affairs* 22(4): 667–86.
Gillespie, M., J. Gow, and A. Hoskins (2007), 'Shifting Securities: News Cultures Before and Beyond the Iraq Crisis 2003: Full Research Report', ESRC End of Award Report, RES-223-25-0063. Swindon: ESRC.
Griffin, L. J. (2007), 'Historical Sociology, Narrative and Event-Structure Analysis: Fifteen Years Later', *Sociologica* 3/2007, doi: 10.2383/25956.
ISR (2011), 'Theory vs. Practice: A Symposium', *International Studies Review* 13 (1 March): 1.
Kahane, H. and N. Cavender (2006), *Logic and Contemporary Rhetoric: The Use of Reason in Everyday Life*, 10th edition (Belmont, CA: Thomson Wadsworth).
King, G., R. Keohane, and S. Verba (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ: Princeton University Press).
Kumar, R. (2005), *Research Methodology: A Step-by-Step Guide for Beginners*, 2nd revised edition (London: Sage Publications Ltd).
Laitin, D. (2002), 'Comparative Politics: The State of the Subdiscipline', in I. Katznelson and H. V. Milner (eds), *Political Science: The State of the Discipline* (New York: W. W. Norton), 630–59.
Laitin, D. (2005), 'Contribution to "Symposium I: Ian Shapiro's The Flight from Reality in the Human Sciences (Princeton University Press, 2005)"', *Qualitative Methods: Newsletter of the American Political Science Association Organized Section on Qualitative Methods* 3(2)(Fall): 13–5.
Luskin, R., J. Fishkin, and R. Jowell (2002), 'Considered Opinions: Deliberative Polling in Britain', *British Journal of Political Science* 32: 455–87.
Macedo, Stephen (2005), *Democracy at Risk: How Political Choices Undermine Democracy and What We Can Do About It*. Washington: Brookings Institution Press.
Martin, L. (1994), *Coercive Cooperation, Explaining Multilateral Economic Sanctions*. Princeton, NJ: Princeton University Press.
O'Flynn, I. (2006), *Deliberative Democracy and Divided Societies* (Edinburgh: Edinburgh University Press).
Parkinson, J. (2006), *Deliberating in the Real World: Problems of Legitimacy in Deliberative Democracy* (Oxford: Oxford University Press).
Pollins, B. M. (2007), 'Beyond Logical Positivism: Reframing King, Keohane, and Verba', in R. N. Lebow and M. Lichbach (eds), *Theory and Evidence in Comparative Politics and International Relations* (New York: Palgrave Macmillan), 87–106.
Putnam, Robert D. (2003), 'APSA Presidential Address: The Public Role of Political Science', *PS: Political Science and Politics* 1 (June): 249–55.
Shapiro, I. (2005a), *The Flight from Reality in the Human Sciences* (Princeton, NJ: Princeton University Press).
Shapiro, I. (2005b), 'A Response to Mackie, Hochschild, and Laitin', *Qualitative Methods: Newsletter of the American Political Science Association Organized Section on Qualitative Methods* 3(2) (Fall): 16–19.
Shively, W. P. (1989), *The Craft of Political Research*, 2nd edition (Englewood Cliffs, NJ: Prentice-Hall, Inc.).
Smith, R. (2004), 'Reconnecting Political Theory to Empirical Inquiry, or, A Return to the Cave?', in E. D. Mansfield and R. Sisson (eds), *The Evolution of Political Knowledge: Theory and Inquiry in American Politics* (Columbus, OH: The Ohio State University Press), 60–88.
Snape, D. and L. Spencer (2006), 'The Foundations of Qualitative Research', in J. Ritchie and J. Lewis (eds), *Qualitative Research Practice: A Guide for Social Science Students and Researchers* (Thousand Oaks, CA: Sage Publications).

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

Part 1

Philosophy of Social Science: Knowledge and Knowing in Social Science Research

2 **Forms of Knowledge: Laws, Explanation, and Interpretation in the Study of the Social World** 27

3 **Objectivity and Values** 64

Forms of Knowledge: Laws, Explanation, and Interpretation in the Study of the Social World

Chapter Summary

This chapter considers fundamental questions about the nature of knowledge and how we acquire it. The key question for political researchers is how we can develop knowledge about the *social world*. A number of answers have been given to this question. Each entails different assumptions about the nature of human behaviour and, consequently, about the methods that are appropriate to investigating and explaining that behaviour. Without some understanding of these questions and the various answers that scholars have made to them it will be difficult for you to design or structure an approach to research into political phenomena, or to make any claim with respect to the findings that result from that research.

We consider three different answers to these questions: those offered by positivism, scientific realism, and interpretivism. In exploring the differences among these approaches and their implications for conducting political research, our discussion will engage with a number of questions, including the following:

- What form(s) of knowledge should be the goal of political research?
- Should the social sciences strive to emulate natural science methods, or is understanding social phenomena something that is essentially different from explanation in the natural sciences?
- Can we study politics scientifically? What does it means to be 'scientific'?
- What distinguishes science from non-science?

Introduction

This chapter begins discussion of a fundamental question of all research: How do we know something is true? To pursue research on social phenomena that is self-aware, systematic, and rigorous, researchers must understand something about the different answers that have been given to this question. There is considerable consensus about the means by which we can acquire scientific knowledge and establish the 'truth' about *natural* phenomena. But the

question of how we go about gaining knowledge of the *social* world is far from settled. At the heart of the debate on this question is whether the methods that scientists have developed to gain knowledge of the natural world can be used to produce reliable and objective knowledge of the social world. Can researchers using these methods establish the truth about some aspect of social life with as much certainty as scientists can about aspects of the natural world?

Some say yes, some say no. The question bears directly on research practice and so continues to be the focus of considerable debate. What do *you* think? It will be interesting to see whether and how your view on the issue changes by the end of this chapter.

In what follows, we consider three different approaches to answering these questions. The one that you accept will determine the sort of research you pursue, the claims you make on the basis of that research, and your assessment of the findings produced by others in our field. Each approach informs a different set of research practices because each is based on different ontological, epistemological, and methodological premises. They consequently differ in their conception of the nature (1) of the social world (ontology), (2) of what sort of knowledge it is possible for us to acquire about this world (epistemology), and (3) of how we go about acquiring this knowledge (methodology). These differences are summarized in Box 2.9 in the concluding section of this chapter.

As indicated, above, the terms 'ontology', 'epistemology', and 'methodology' relate to fundamental issues concerning research practice and knowledge. Ontology is concerned with '*what is*'. Here, we are concerned with the nature of the social world and the basic elements that make up this world. In political research, ontological questions include whether the social world is fundamentally different from the natural world; whether it is an objective reality that exists independently of us or is in important respects subjectively created. Epistemology is concerned with *what is knowable*. What can we know about social phenomena? What type or form of knowledge can we treat as legitimate knowledge about the social world? **Methodology** is concerned with *how we obtain knowledge*. What are the means and methods that can provide us with legitimate knowledge of the political world? Box 2.1 shows how these key issues concerning knowledge are related.

Recall the question we posed previously: Can the methods developed by scientists to gain knowledge of the natural world enable us to produce reliable and objective knowledge of the social world?

We begin this chapter by considering the arguments of those who answer 'yes' to this question. This answer has been elaborated in two traditions of thought: in positivism, and in scientific realism. While these traditions, or 'philosophies of social science', differ from each other (scientific realism challenges key positivist assumptions), both claim that the

BOX 2.1 Ontology ▶ Epistemology ▶ Methodology

Ontology	Epistemology	Methodology
What exists? What is the nature of the social world?	What sort of knowledge of it is possible? How can we know about it?	What strategies can we use to gain that knowledge?

social world is no different than the natural world and so can be studied the same way that scientists study the natural world. To be clear, the claim here is that we can use the same methods used to study natural phenomena in order to establish, with the same degree of certainty, the truth about any social phenomena we wish to understand.

After considering the 'yes' answer to our question, we then consider a third approach, interpretivism, whose adherents argue that the social world is *not* the same as the natural world. They argue that, while the natural world exists independently of the person scientifically observing it, this is not true of the social world; and that when the tools scientists employ to study natural phenomena are used to study social phenomena, they do not enable us to achieve the objectivity and scientific certainty that they permit when used to study the natural world.

The three approaches outlined above differ from one another with respect to the question of whether scientific methods can be used to study human behaviour. From this difference flow additional differences which, taken together, generate different types of research and research practices. As you shall see, each approach has developed, in part, through a critique of the other. However, each produces useful forms of knowledge and, thus, contributes to the range and type of questions that political research can effectively address.

Positivism

Before discussing the basic tenets of positivism and tracing its development, it is worth noting how important this system of thought has been to the development of political research.

Positivism began as a movement to establish a sound basis for social scientific inquiry. This is a key issue in political research. Political researchers want to be able to offer credible answers to important questions, and they are concerned to ensure that the research practices and methods they employ enable them to do this. Positivism offers one approach to resolving this issue: it maintains that researchers can arrive at factual, reliable, and objective answers to questions about the social world by employing the methods used in the natural sciences. Some researchers see this claim as highly controversial; others consider it to be plain common sense. Consequently, positivism plays a dual role in our field: for many, it provides a foundation for research; for others, it stimulates the articulation of alternative methodological positions.

Positivism began to transform political research beginning in the 1950s and early 1960s. In what scholars in our field have called the 'Behavioural Revolution', researchers turned to a more social scientific and empirical approach to studying political behaviour. **Behaviouralism**, the term used for the application of positivist tenets to political research, is concerned with the question of what political actors do and why they do it. The key tenet of behaviouralism is that only observable behaviour may be studied. Thus, for behaviouralists, political research involves studying and explaining the observable behaviour of individuals or aggregates of individuals (parties, classes, interest groups, governments, social movements). Until the mid-1970s, behaviouralists tended to focus on questions that could be answered by gathering and studying data conducive to exact measurement, as for instance voting data or data from public opinion polls and social surveys. But critics argued that, by focusing on phenomena that lent themselves more easily to measurement, the field

had become preoccupied with technique rather than substance, and was failing to address significant problems. It was this concern that triggered a 'post-behavioural revolution'.

This 'revolution' was not concerned to displace behaviouralism. Its aim, as a key figure in both the 'Behavioural' and 'Post-Behavioural' revolutions put it, was to 'propel political science in new directions' (Easton 1969: 1051). Some of these new directions moved the field towards a further realization of positivist and behaviouralist goals, such as the trend in political research towards 'positive political theory' or **rational choice theory**. Positive political theory assumes that what motivates behaviour is rational self-interest, 'as opposed to attitudes, which are the subject of study in much behavioural research', and that individual self-interested rational action combines to produce collective political outcomes (Amadae and Bueno de Mesquita 1999: 270). But the post-behaviouralist revolution also set in motion trends that moved the field in non-positivist directions. These directions included the rebirth of normative theory, which we will consider in Chapters 3 and 5; and the emergence of a set of approaches associated with 'interpretivism', including constructivism, feminism, post-modernism, and critical theory, which we will discuss further along in this chapter.

The behavioural revolution changed our field in a number of important ways. It opened up an important discussion and debate about the desirability and possibility of using scientific methods to attain reliable, empirical, causal knowledge about political life. While originally a too narrow focus on observable and measurable phenomena meant that too much emphasis was placed on what might be characterized as a narrow range of questions, behavioural researchers also succeeded in broadening the domain of political research by drawing on theories, research methods, and findings from other disciplines for insights into political behaviour. Behaviouralism established an emphasis on research based on empirical observation, testing involving systematic evidence, and falsifiable and causal explanation. Its emphasis on the importance of research that is capable of replication by others, made researchers more precise about what they want to know, what explanation they are advancing, and how they intend to demonstrate it.

We have gained a sense of how, through behaviouralism, the application of positivism to political research has shaped the values, beliefs, and rules that have guided analysis within our field. We turn now to a discussion of its basic tenets. In discussing these tenets, we first focus on the classical positivist tradition, and then explore further developments in positivist thought. Our aim here is *not* to provide an intellectual history of positivism, but to provide an understanding of the array of positions that today in our field are associated with positivist thought.

Classical positivism

It is interesting to consider how the term 'positivism' arose. The term was originally coined by the French philosopher Auguste Comte (1798–1857) as part of an argument about how human societies had gone about searching for the truth. According to this argument, human society has experienced three phases in the course of its search for the truth. Society first developed religions: systems of thought that offered a guide to making sense of the world. Following this theological phase or stage, it then expanded its belief in things that cannot be observed as, for instance, in its development of metaphysics. Now, Comte had argued, it had entered into a final 'positive' stage in which the search for truth is characterized by the

systematic collection of observed facts. The term 'sociology', which refers to the scientific study of the social world, was also his invention. Both terms expressed the same belief: that the social world could be explained using similar methods to those used to explain natural phenomena.

This view of social science methodology, like the others we will discuss in this chapter, commits us to a number of ontological and epistemological claims. The nature and implications of these claims will become clear as we identify the edifice of thought which positivism builds concerning the nature and goals of social scientific inquiry. This edifice rests on four basic tenets. The first of these maintains that there is no difference between the natural and social sciences; and the remaining three tenets develop this comparison.

The *first* tenet of positivism, as we've just indicated, is the claim that there are no fundamental differences between the natural and the social sciences. This claim is what is called '**naturalism**'. Note that naturalism presupposes an ontological claim about the social world: if there is no difference between the social and natural *sciences*, it must be because there is no fundamental difference between the social and natural *worlds*. Now, if the nature of the social world is no different from that of the natural world, then like the natural world, there exists in the social world a mind-independent, objective reality that exists independently of our knowledge of it. This related assumption, which we call 'realism', we will be discussing further on in this chapter.

A *second* tenet of positivism is **empiricism**. Empiricism is a philosophical theory of knowledge which answers an epistemological question: how can we acquire knowledge of the world? The answer, according to empiricism, is that what we know of the world is limited to what can be observed. Knowledge is only that which originates in sensory experience: there is no a priori knowledge, no knowledge of reality that is acquired prior to sense experience. When we adopt a positivist approach to social scientific inquiry, then, we are committing ourselves to the notion that social reality can only be known through what is observed and that knowledge of the social world is therefore limited to phenomena that can be observed though the senses.

In a further elaboration of the first tenet, a *third* tenet of positivism is that the goal of social science is to explain and predict social phenomena by means of **laws**. Positivism maintains that, since the social sciences are no different from the natural sciences, they should have the same structure and logical characteristics as the natural sciences. Thus, as the German logician Carl Gustav Hempel (1905–97) argued, if the discovery of laws is necessary in the physical or natural sciences, then laws must be necessary also in social science. If the social world is like the natural world, then, like the natural world, it also must be regular, systematic, and law-governed. There are regularities in, and ultimately laws of, social and political processes; and we can explain social events and phenomena by means of law-like generalizations that have the same status as natural scientific laws.

The possibility of discovering laws in the social sciences is one of the key issues on which positivism and its critics divide. Do social laws exist, they ask? Those who answer 'no' insist that there is a difference 'in kind' between the subject matter of natural and of social science, which makes it impossible to establish social laws (McIntyre 1994: 131). We will be considering this view later in the chapter.

We have said that positivism holds that the social world is regular, systematic, and law-governed, like the natural world; that social phenomena can be explained and predicted by

means of laws that have the same status as natural scientific laws; and that the purpose of social science, therefore, is to discover these laws. But how do we go about discovering laws of social life? According to classical positivist thought, laws can be discovered through systematic investigation of observable events and happenings, and through a means or process of reasoning called '**induction**'. When we engage in a process of inductive reasoning we begin with specific observations and measures (gathering facts). We then move on to identifying patterns and regularities in these observations and to the formulation of some tentative hypotheses. Finally, we develop some general conclusions or theories. We will have more to say about induction, along with other means of reasoning, further on in our discussion.

We are still discussing the third tenet of positivism: the view that the goal of social scientific inquiry is to explain social phenomena through the discovery of laws. For positivism, the goal of social inquiry should also be to establish causality, or a cause–effect relationship, between events in the world. Positivism sees the social world as comprising phenomena that are causally related to each other; consequently, to explain a social outcome we must show the factors that combined to bring it about or caused it to be more likely to occur in the circumstances.

Virtually all social research is concerned to discover causes. But there are different conceptions of causation. We will be discussing causality at various points in this book. But here, we will start that discussion by considering the positivist conception. This is an empiricist one which was introduced by the Scottish philosopher, economist, and historian, David Hume (1711–76). Most of us probably carry in our minds an idea of causation as expressing some type of 'necessary connection' between two events (the cause and the effect). But Hume pointed out that a 'necessary connection' or causal relationship cannot be directly perceived; 'we cannot in any instance discover a power, necessary connexion, or quality' which renders one event 'an infallible consequence of the other' (1966: 51). Instead, what we observe is only the 'constant conjunction' of events. Recall that, according to empiricism, we can only know what we can observe. What we can observe is that one thing follows the other, but we cannot observe some underlying generative force that makes this happen. Instead, we discover observable regularities; and the experience of observing this 'constant conjunction' between events conveys to our minds a necessary relation between these events. The causal conclusions we reach are based, therefore, not on 'knowledge of causal mechanisms and the generative properties of things', but only 'on the observation of how a certain event is followed again and again by a certain other event' (Ekström 1992: 108).

According to this conception, then, causation is constituted by facts about empirical regularities among observable variables. All we can do is observe that one thing follows another with regularity; and, because of this observation, we develop a psychological expectation that Y will occur whenever X does. But we cannot know that X is the *cause* of Y by observing that X is constantly followed by Y. Consequently, positivists are concerned with observing empirical regularities rather than in discovering causal mechanisms. This conception of causation as a perceived regular association among variables is a *subjective* one. According to positivism, an *objective* conception of causality, one involving causal necessity or causal mechanisms, is metaphysical.

A *fourth* tenet of positivism is that in the pursuit of knowledge of the social world through application of the scientific methods used in the natural sciences *it is possible to distinguish* between *facts and values*, and to treat 'facts' as independent of the observer and

the observer's values. We can gain value-free, objective knowledge of the social world because it is possible for us to distinguish statements of fact (confirmed by the senses) from normative statements. Science, which is concerned with the discovery of facts, states as 'fact' only that which can be confirmed through observation. Scientific investigation can therefore be clearly distinguished from normative inquiry, which is concerned with values and ethics as, for instance, in policy studies. As we shall see, this tenet represents a key difference between positivism and alternative approaches. The various arguments advancing and rejecting this fact-value distinction leads us in some interesting directions, and we will be exploring these in some detail in Chapter 3. Here, however, we turn to a consideration of developments which extended and modified classical positivist thought.

Empiricism and logic as the basis of truth claims

Classical positivist tenets have been the focus of considerable debate, both among positivists and non-positivists. We will be considering the critiques that have been offered by non-positivists further on in this chapter. Here, however, we focus on two key critiques by positivists which have had a significant impact on positivist thought. The first of these critiques came from a movement which adopted the name 'logical positivism'. Logical positivism succeeded in introducing an important extension of positivist thought. But then, logical positivism itself became the target of a critique by the philosopher of science, Karl Popper, and this highly influential critique brought about an important further development within positivism.

Logical positivism began in the early twentieth century as a movement within philosophy. Positivism, you will recall, maintains that we can acquire knowledge of the world only through empiricism—that what we know of the world, and what we can therefore treat as legitimate knowledge of the world, is limited to what can be observed. But logical positivists argue that logical reasoning and mathematics should also be treated as sources of knowledge, in addition to empiricism; that knowledge of the social world comes, not just through what can be observed (empiricism), but also through logical reasoning. Following on from the idea that knowledge of the social world comes through what can be observed (empiricism) and also logical reasoning, logical positivism maintains that social inquiry should therefore combine induction (based on empiricism) and deduction (in the form of logic) as methods of reasoning.

We have already briefly discussed induction as a method of reasoning. As Figure 2.1 shows, induction begins with particular observations or cases and then develops generalizations about them. If we accept the classical positivist tenet that empiricism is the only means of gaining legitimate knowledge of the social world, than our process of reasoning should begin with specific observations and measures (gathering facts). However, if we accept the logical positivist argument that knowledge can be gained, not only through observation, but through logical reasoning then, in addition to inductive reasoning, we would want to employ a second method of reasoning: deduction. **Deduction**, as Figure 2.2 shows, moves not from specific observations to broader generalizations and theories (as in inductive reasoning), but the other way around: it begins with broad generalizations or theoretical propositions and then moves to specific observations. In using deduction, then, we start, not with an observation, but either with a theory that has already been confirmed or with a logical argument, and then we draw out the meaning or implications this has for explaining some particular case or phenomenon. This tenet of logical positivism—that social inquiry

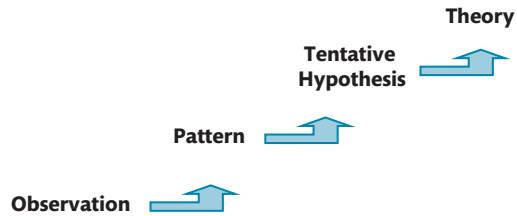


Figure 2.1 Induction

Source: Reproduced with permission from W. Trochim et al. *Foundations of Research Methods. In Research Methods: The Essential Knowledge Base,* 2nd Edition. Copyright © 2016, Cengage Learning.

should combine induction (based on empiricism) and deduction (in the form of logic) as methods of reasoning—represents a substantial development of positivist thought.

Before moving on let's consider an example of how each of these two processes of reasoning works. Box 2.2 provides an example of how inductive reasoning proceeds.

BOX 2.2 Inductive Reasoning

In an article entitled 'The Natural Birth of a Hypothesis' (1963), a US researcher, Milton Leitenberg, made a visit to a severely depressed area of Nova Scotia. One evening he sat watching a group of children playing what appeared to be the familiar game of 'Tag'. However, as he continued observing the children at play, he realized that they were playing the game quite differently than it was played in the United States. It appeared that the goal of the game as these children were playing it was *not* to catch the person who was 'it': instead, what he observed was that those who were seemingly pursuing 'it' slowed down when 'it' slowed down, and they sped up when 'it' sped up. Moreover, 'it' seemed not to make any attempt to outrun the pursuers. As he puzzled over what he had observed, Leitenberg wondered if there was a connection between this cooperative version of the game 'tag' and the depressed economic situation of the community. He hypothesized that this form of the game was linked to the children's experiences of economic hardship (and a consequent lack of children's toys and recreational facilities). He then speculated, more generally, on how social and economic changes affect children.

Here, then, is a process of reasoning in which a theorist starts with a particular observation and then, based on that observation, develops a generalization about the impact of economic change on children.

Deductive reasoning works the other way around: rather than starting with an observation, deduction starts with a theory that has already been confirmed, or with a logical argument, and then draws out the meaning or implications it has for explaining some particular case or phenomena. Box 2.3 provides an example of how this works.

We have now identified two different methods of reasoning. It should be noted, however, that in practice researchers do not use solely one method or the other. Scientific inquiry typically involves *a process of continuous interaction between theory and observation*, in which the researcher moves from observation to theory (induction) and from theory back to observation (deduction). Box 2.4 illustrates how this process contrasts and combines induction and deduction. The compiling of evidence (induction) leads the researcher to theory (deduction); and once a hypothesis is formed, the researcher brings it 'backward' for readjustment or redefinition. The term '**retroduction**' describes this interaction of induction and deduction in an evolving, dynamic process of discovery and hypothesis formation.

BOX 2.3 Deductive Reasoning

With deduction, a theorist might start with a line of reasoning that because candidates in elections want to win, they will try to appeal to more of the electorate by appearing to moderate their positions. It is logical, therefore, to expect that as an election campaign progresses, candidates will move more towards the centre of the political spectrum. From this general statement, we could then hypothesize that, as we get closer to a general election, the main candidates will moderate their stated positions. We would conclude this, not on the basis of a detailed examination of their personalities or backgrounds, or on the basis of an investigation of a large number of previous electoral campaigns, but on the basis of a general logical proposition. After the election is concluded, we could then test the hypothesis by studying campaign speeches during the last months of the electoral cycle with those from an earlier point in the cycle. We might then further test the hypothesis by examining the campaign speeches of candidates in former elections. If our analyses confirm the hypothesis, we have a theory of electoral politics.

How we might experience this in the research process is what one of us describes as 'mucking about': we muck about in the dust of detail until some theoretical generalization or relationship comes to mind that seems to offer an explanation of what our empirical research has revealed; we then work to develop this theoretical notion until we hit a brick wall and then we go back to sifting through our empirical research until some additional inspiration enables us to return to, and further refine, our theoretical ideas.

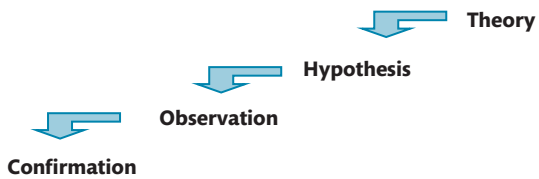


Figure 2.2 Deduction

Source: Reproduced with permission from W. Trochim et al. *Foundations of Research Methods. In Research Methods: The Essential Knowledge Base,* 2nd Edition. Copyright © 2016, Cengage Learning.

Logical positivism made two contributions to positivist thought. Its first contribution, as we have just discussed, was to introduce the idea that social inquiry should combine both induction and deduction. A second contribution, which we will discuss below, was to establish 'verification' (of statements or propositions) as the criterion for establishing truth claims, and as a means of defining a clear line of division between science and metaphysics.

Both of these logical positivist positions became the target of a critique by Karl Popper (1902–94), a philosopher of science who also wrote extensively on social and political

BOX 2.4 Induction, Deduction, Retroduction

Induction	Observation → theory
Deduction	Theory → observation
Retroduction	Observation ↔ theory

philosophy. Popper's critique had a decisive impact on social scientific thought. In fact, its influence was so great that logical positivism's most important contribution to social science, it might be argued, is the role it played in having served as the focus of this critique. This does not diminish its contribution: in the quest to establish a sound basis for scientific inquiry, logical positivism raised important questions about the concepts and practices of science which continue to have relevance for social scientific inquiry today. Moreover, while Popper was a critic of logical positivism, there are also many affinities between his views and those held by logical positivists.

Let's then turn to Popper's critique of Logical Positivism. As we have discussed, logical positivists contributed to positivist thought the argument that both inductive and deductive methods of reasoning should be used to acquire knowledge of social phenomena. Popper, as we shall see, *rejects induction entirely* as a method of reasoning for explaining social phenomena. Moreover, his argument for rejecting induction entails a rejection, as well, of the logical positivist emphasis on **verifiability** as a basis for establishing truth claims.

So, why does Popper reject the idea that we can use induction as a method of reasoning? He builds his argument against inductive reasoning by first addressing what David Hume calls 'the problem of induction'. This is the problem of whether observation can provide the basis for gaining general theoretical knowledge. Experience, as Hume points out, is particular. So the question is, how can we achieve knowledge—which is general or even universal, and how can we reach general statements of scientific law, on the basis of experiences that are necessarily limited and particular? Popper argues that no matter how many experiences we have of observing something, this does not permit the deduction of a general statement of scientific knowledge.

The reasoning that leads him to this conclusion begins with David Hume's argument about the limits of inductive reasoning. Hume argued that since we cannot observe the universe at all times and in all places, but are only able to observe particulars, we are not justified in deducing general laws based on inductive evidence. Popper's now famous story of the black swan illustrates what happens when we attempt to formulate laws based on observation. The story is that, once upon a time, Europeans thought all swans were white because, having found nothing but white swans for thousands of years they concluded on the basis of their experience that all swans were white. But one day Europeans went to New Zealand (as Popper had), and there they found black swans. What this story tells us is that no matter how many observations confirm a theory, it only takes one counter-observation to falsify it: only one black swan is needed to repudiate the theory that all swans are white. And since it only takes a single unforeseen or seemingly improbable event to invalidate a generalization based on empirical observation, then empirical observation alone cannot generate 'laws'. Popper therefore concludes that, rather than endeavouring to discover laws through induction, what scientists should be doing is testing theory deductively.

Popper's argument for rejecting induction leads him to reject, as well the second contribution logical positivism sought to make to positivist thought: the idea that scientists should seek to verify hypotheses. If a single exception to the rule destroys inductively generated theory, then conclusive verification of a hypothesis, he argues, is not possible. Popper proposes that to establish truth claims we should in fact do the *reverse* of what logical positivists propose: rather than continually attempting to *verify* or *prove* a theory, scientists should attempt to *disprove* it. Since, as the 'problem of induction' demonstrates, we cannot

verify a **hypothesis**, our aim should be to *falsify* it. What this requires is that we formulate propositions in such a way as to enable them to be refuted. By doing this, it will be possible for us to show a theory to be wrong; and we can then introduce a new theory which better explains the phenomenon. This is how we achieve progress in science, Popper argues. But, for Popper, **falsifiability**, rather than verifiability, not only represents the way forward in science; it defines the boundary between science and pseudo-science or metaphysics. Anything non-falsifiable, he argues is outside science.

Let's think about this claim about falsifiability with regard, for instance, to religion and ideology. Religions and ideologies provide a guide for understanding the world. They are similar in that a characteristic of both is that they are non-falsifiable: potentially disconfirming or anomalous facts do not prove them false, but are incorporated within them. In contrast, the requirement that a scientific theory must state what evidence would disconfirm it or prove it to be false means that, if you cannot think of anything that might disconfirm a theory, then it is not a theory at all but a set of self-verifying statements—an ideology.

So what we have seen is that, in rejecting induction, Popper was rejecting the idea that observation provides the basis for the formulation of scientific theories. Theories cannot be derived from observation (induction), because at any time a single observation can disconfirm the theory. Popper concludes that social inquiry must proceed deductively, through a process in which we use observations, not as the basis of theories, but in order to 'test', or falsify, them. Our endeavour must be to falsify hypotheses, to reject those which are falsified, and to continue to test those that are not until they become so thoroughly tested that we can consider them to be 'confirmed', though it remains possible that some day someone may falsify or significantly modify them.

There are a number of objections that have been made to these arguments. Two of them will be discussed in Chapter 3; but it is worth flagging them up here. The first is that Popper seems to define a distinction between facts and theories, to assume that the observations or facts that we pursue as a means of testing theories can be established independently of the theory that they are meant to test. As we shall see in Chapter 3, there is an important debate about this issue. The second objection, as we shall see in Chapter 3, is that Popper's notion of falsifiability is at odds with how scientists actually go about developing and testing theories. Do researchers *really* seek to disprove or falsify their own theories? Do they discard their theories when they are confronted with disconfirming evidence? We will consider two answers to these questions, and the further evolution of Popper's falsificationist position that developed as a response to them, in Chapter 3.

Before moving on to a consideration of non-positivist approaches, there is a question that remains to be addressed concerning Popper's argument that deduction should be the sole basis for explanation. To briefly sum up before moving on to consideration of this question: *classical positivism* maintains that we can discover laws of social life that explain social phenomena by using inductive reasoning based on systematic investigation of observable events and happenings. *Logical positivists*, however, argue that both induction, based on empiricism, and deduction in the form of logic can be used to discover laws. However, the implication of *Popper's critique* of both classical and logical positivism is that it is *only through deduction* that we can establish laws of social life as a basis for explanation. But this raises the question of how we use deductive reasoning to discover laws of social life as a basis for explanation. What is the process through which deduction operates as a

means of explaining social phenomena? For the answer to this question, we turn again to Carl Gustav Hempel.

Hempel (1994) develops a model of how deductive reasoning as a means of discovering laws constitutes or works as an explanation for things we observe in the social world. The model is based on a number of assumptions, some of which will be familiar to us by now. Hempel assumes that explanation in the social sciences, as in the natural sciences, involves the search for and discovery of law-like generalizations; and that the logic and function of what he calls 'general laws' are the same in both the social and the natural worlds. Thus, in both the natural and social sciences, individual events can be explained by reference to general laws of nature: to explain some event is to cite some law or laws and other conditions from which the event can be deduced.

Hempel (1994) formalizes this definition of explanation in his **deductive-nomological model**. A deductive-nomological explanation is deductive because we can logically deduce the phenomenon we want to explain (*explanandum*) from that which does the explaining (the *explanans*); and it is nomological because the *explanans* includes at least one law ('nomos' is the Greek word for law). According to this model, then, something is explained when it is shown to be a member of a more general class of things, when it is deduced from a general law or set of laws. For instance,

> To explain fully why an actor votes (a 'fact') we must do more than just isolate the particular cause of this particular event (for example, the intensity of the voter's concern about unemployment). We must subsume this act of participation under a 'law' that explains why, under certain conditions, the voter had to vote: 'persons with intense preferences for candidates or issues', everything else being equal, will become 'active in politics'.
>
> (Milbrath 1965: 53; quoted in Schwartz 1984: 1123)

Given the general law, the particular case in question was to be expected.

But how do we know that what we take to be a 'law' actually *is* a law? A regularity that we observe might either be a true, law-like generalization, or only 'accidentally' true as a result of circumstance or coincidence. How can we confirm that what *appears* to be a law actually *is* a law? The question of how to distinguish generalizations that are law-like from ones that are 'accidental' is a central problem in the philosophy of science. Hempel (1994) argues that, in general, we can say that a law expresses a necessary connection between properties, while an accidental generalization does not; and that if a necessary connection exists between its properties, then we should be able to test a law by its ability to predict events. If we predict something on the basis of a law and find that the prediction was true, then the law can be said to be confirmed.

These assumptions are the basis for the model that Hempel develops to describe how we can confirm that what appears to be a law actually is a law. According to his **hypothetico-deductive model** of confirmation, we confirm that the generalization is a law (rather than an accidental generalization) by treating it as a hypothesis, and then we test the hypothesis by deducing from it a sufficient number of explicit predictions of further phenomena that should be observable as a consequence of the hypothesis. Observations that run contrary to those predicted are taken to conclusively falsify the hypothesis; observations which are in agreement with those predicted are taken as corroborating the hypothesis. It is then supposedly possible to compare the explanatory value of competing hypotheses by looking to see how well they are sustained by their predictions.

Here is an example of how this works. What in our field is called Duverger's Law is regarded as a law, or at least a law-like generalization. Maurice Duverger proposed that a

plurality rule election system, one in which the winner of an election is the candidate who wins a plurality, rather than a majority, of votes, tends to favour a two-party system. There are two theoretical explanations for why this is. The first is the 'mechanical effect' of under-representing losing parties; and the second is a 'psychological factor': voters don't want to waste their votes on losers (Riker 1982: 761). William Riker explains: 'when the definition of winning forces candidates to maximize votes in order to win (as in plurality systems), they have strong motives to create a two-party system; but when the definition of winning does not require them to maximize votes (as in runoff and proportional systems), then this motive for two parties is absent' (Riker 1982: 755).[1]

To sum up: the deductive-nomological model holds that an observed phenomenon is explained if it can be deduced from a law-like generalization. The hypothetico-deductive model confirms that a generalization is a law by treating the generalization as a hypothesis, and testing it by its deductive consequences. To explain some fact is to cite a law or laws plus other relevant conditions from which the *explanandum* may be deduced (the deductive-nomological model of explanation). To confirm a hypothesis is to deduce some observed phenomenon from the hypothesis plus other relevant known conditions (the hypothetico-deductive model of confirmation).

Before moving on, let's review some of the terms that have been introduced in this chapter thus far (Box 2.5).

BOX 2.5 Key Terms

Induction	A process of reasoning that moves from specific observations to broader generalizations and theories.
Deduction	A process of reasoning that begins with broad generalisations or theoretical propositions and then moves to specific observations.
Retroduction	The interaction of induction and deduction in an evolving, dynamic process of discovery and hypothesis formation.
Verifiability	The claim that the criterion for establishing truth claims, and thus the goal of social scientific inquiry, should be to verify statements or propositions.
Falsifiability	The claim that (1) a scientific theory must be formulated in a way that enables it to be disconfirmed or proven false; if it can't be falsified it is not a theory but an ideology; and (2) when we 'test' a theory we should seek, not to verify, but to *falsify* it.
Deductive-nomological model	According to this model, something is explained when it is shown to be a member of a more general class of things, when it is *deduced* from a general law (*nomos*, in Greek) or set of laws.
Hypothetico-deductive model	According to the model, we confirm that a generalization is a law by treating it as a hypothesis, and then we test the hypothesis by deducing from it predictions of further phenomena that should be observable as a consequence of the hypothesis.
Causal mechanism	Something that links a cause to its effect, that generates some type of 'necessary connection' between two events (the cause and the effect).

We have traced the development of positivist thought through a consideration of the basic tenets of classical and logical positivism and the arguments advanced by Karl Popper. We turn now to approaches that emerged as a challenge to positivism and positivist research.

Challenges to positivist approaches in the social sciences

There are a number of approaches to social inquiry that challenge the positivist position and that articulate a fundamentally different basis for inquiry. Here, we focus on two alternative positions—those represented by scientific realism and interpretivism.

Scientific realism

Scientific realism is similar in some ways to positivism because it accepts some key tenets of positivism.

It accepts postivism's 'naturalist' ontology which assumes that the social world is not different from the natural world. It assumes that both worlds are governed by laws and that, in both, there is an objective reality that exists independently of our knowledge of it. However, while it accepts this tenet of positivism, it breaks distinctively with a second tenet of positivism: empiricism. You will recall from our discussion, above, that empiricism is a philosophical theory of knowledge which claims that what we know of the world is limited to what can be observed. However, scientific realism assumes that what is objectively 'real'—both in the natural and the social worlds—consists, not only of what we can directly observe, but also of *unobservable* elements: elements, that while unobservable, operate to produce real outcomes and that should, therefore, be treated as real.

Given an ontology that assumes that what is 'real' consists of both observable and unobservable elements, scientific realism cannot then depend on an *epistemology* that maintains that only that which we can directly observe can be known. Thus, scientific realists maintain that explanation can be based, not only on an observable, law-like regularity, but can include, as well, an unobservable **causal mechanism** that links a cause to its effect.

We discussed the positivist assumption that statements that are not based on observable data are metaphysical. Scientific realists reject this assumption. They assume that there are knowable, mind-independent facts, objects, or properties that cannot be directly observed but which are, nonetheless, *real*. They argue that there are unobservable elements of social life, such as structural relations between social phenomena, that these unobservable elements are crucial to an understanding and explanation of what goes on in the world, and that this is precisely what makes the social world similar to the natural world. This ontological conception of the social world is, therefore, not metaphysical, but more closely aligned with the tenets of the natural sciences, than the positivist conception. In fact, scientific realists claim, this non-positivist version of science is *more* scientific than positivism. Their message, Ruth Lane notes (1996: 373), is that 'we don't have to be positivists to be scientific!'.

Consequently, for scientific realists, the goal of scientific research is to describe and explain both observable and unobservable aspects of the world. But how can we establish the reality of things that cannot be directly observed? How do we know these unobservable elements exist? According to scientific realism, we know that unobservable elements of social life exist because we can observe their *consequences*. If unobservable elements of social life produce observable effects then they can be treated as 'real'. To posit the existence of unobservable entities to explain observable outcomes is consistent with well-established scientific practice. For instance, we treat gravity and subatomic particles as real because, even though we cannot see them, we can see their effects. Similarly, in social and political life there are many elements that are not directly observable—social structures, capitalism, society—but that, nonetheless, have observable effects.

For scientific realists, then, scientific knowledge does not take the form solely of empirical regularities. It therefore cannot be *solely* concerned with the goal of formulating law-like generalizations based on observations. If scientific realists reject an *ontology* that assumes that only entities of which we have direct sensory experience are 'real', then they must also reject an *epistemology* that assumes that knowledge of the social world can only be gained through direct observation. Nor can they accept the empiricist (Humean) conception of causality that positivists employ.

Recall that positivists treat causation as constituted by empirical regularities among observable variables, and seek to establish causal relationships by observing these regularities rather than by discovering causal mechanisms. Since we are unable to have knowledge of causal mechanisms through direct observation, they consider the notion that causal mechanisms produce outcomes in social life to be metaphysical. But scientific realists assume that *both* observable and unobservable elements exist in social life. Consequently, their conception of causality includes both regularities among observable elements as well as unobservable causal mechanisms that generate those regularities. For scientific realists causal powers are 'real'; they are a legitimate object of scientific investigation, and fundamental to explanations of social outcomes. A scientific realist explanation of social outcomes therefore entails providing an account of the causal mechanism that brought about a given outcome; and developing empirically justified theories and hypotheses about causal mechanisms.

A causal mechanism can be defined as 'the pathway or process by which an effect is produced or a purpose is accomplished' (Gerring 2007: 178). Charles Tilly identifies three sorts of causal mechanisms that operate in the social world: environmental, cognitive, and relational. Environmental mechanisms are 'externally generated influences on conditions affecting social life'; cognitive mechanisms 'operate through alterations of individual and collective perception'; and relational mechanisms 'alter connections among people, groups, and interpersonal networks' (2001: 24). An example of an environmental mechanism is one that has been used to explain the apparent link between oil exports and authoritarian rule: the 'rentier effect' generated in oil-exporting countries by the government's reliance on oil revenues. Michael Ross describes this 'rentier effect' as follows: 'when governments derive sufficient revenues from the sale of oil, they are likely to tax their populations less heavily or not at all, and the public in turn will be less likely to demand accountability from—and representation in—their government' (2001: 332). Cognitive mechanisms—changing conceptions of racial, ethnic, gender, religious, or class differences—have been identified to explain

ethnic conflict (e.g. Hoffmann 2006). Relational mechanisms, such as the bureaucratic containment of previously autonomous military forces, have been held to affect 'the likelihood of civil war' (Tilly 2001: 38).

The idea that such unobservable 'social mechanisms' can be used to explain outcomes in the social world has generated considerable debate. What is the scientific status of unobservable entities? Debate on this issue focuses on a number of questions:

1. What is the ontological status of macro-social structural or institutional mechanisms used to explain social outcomes?
2. How do we explain macro-social phenomena?
3. How do macro-social 'social mechanisms' produce social outcomes?

1. What is the ontological status of macro-social mechanisms used to explain social outcomes?

The first of these questions is an ontological one: what are the basic entities that make up the social world. Our answer to this question determines our choice of what to treat as the basic unit of analysis in social inquiry. Positivist and behavioural research treats individuals as the basic unit of social analysis and consequently favour individual-level mechanisms to explain outcomes. So, for instance, rational choice approaches, which seek to explain social phenomena as the result of the purposive choices by individuals within a given set of material and structural circumstances, assume that the instrumental rationality of individuals is the causal mechanism that produces social outcomes. In other research, the unit of analysis is some larger macro-social entity or collective whole. For instance, research in international relations that employs a world systems approach seeks to explain social outcomes as a result of structural or institutional mechanisms related to the operation of the world capitalist system. But debates about unobservable causal mechanisms tend to focus, in particular, on the nature of macro-social and collective entities and their role in explaining social outcomes, and to revolve around two broad positions on these ontological and methodological questions. These positions are shown in Box 2.6.

BOX 2.6 Individualism and Holism

	Individualism	Holism
Ontology	The basic units of society are individuals, and social phenomena are the combined results of individual actions.	The whole of something is distinct from and not directly explicable in terms of its parts. Social facts have social causes that are irreducible to facts about individuals.
Methodology	Explanations of social phenomena—classes, power, nations, or other social phenomena—must be reducible to the characteristics of individuals.	The properties of a system as a whole cannot be deduced by the properties of its components alone. The system as a whole determines how the parts behave. Each social entity (group, institution, society) has a totality that is distinct.

2. How do we explain macro-social phenomena?

Political researchers continually refer to unobservable macro-social entities or social collectives such as 'society' and 'structures of power' or nations, to explain political events and processes. But what sorts of things are we investigating when we do research on, for example, the capitalist system or globalization, or on nations, political parties, or classes? Are social 'wholes' or collectives such as 'society' or 'the working class' reducible to the individuals who make them up (**individualism**)? Or are they more than the sum of their individual constituents and actors in their own right (**holism**)? Should social explanation give primacy to individuals (**methodological individualism**) or to social structures, institutions, and collectives (**methodological holism**)?

According to holism, social life is not simply a sum of individual action. Much human interaction consists of actions, and generates outcomes, that cannot be comprehended and explained in this way because the interaction of individuals produces 'novel' properties—properties that are not possessed by the individuals taken in isolation. In the same way, we can say that society in certain of its aspects is not reducible to the individuals that comprise it, even though these aspects of society may be rooted in individual practice. Interactive relations between individuals can produce properties that are separate and distinguishable from the individuals themselves. Consequently, an individualistic form of methodology may likely obscure just what a researcher might want to learn.

However, approaches that place social structures at the forefront of explanations of social phenomena sometime tend to imply a causal determinism in which the role of human agency is lost. In these approaches, structures seem to exist separately from, but nevertheless to determine, individual social action. This leads to the tendency to treat these macro-social structural entities as if they had a concrete, material existence; to treat them as analytically independent of their constituent elements; inert, unchanging, and unmediated by human agency. This is what we call; 'reification'.

The 'reification' of macro-social entities or structures is apparent, for instance, in many discussions of globalization. Some analysts trace the emergence of globalization to the interests and activities of specific groups and governments. But the predominant tendency is to locate the origins and driving force of globalization in macro-sociological entities like markets and capital. Globalization, in this view, is the outcome of the 'logic' of markets or of international capital, or of institutional and technological innovation and change, or of the evolutionary development of capitalism and capitalist production. Globalization appears, from this perspective, as a more-or-less natural outcome of a natural spatio-temporal spread of more-or-less natural economic and technological developments. 'Markets', 'technology', 'capital', and 'capitalism' are treated as primary and autonomous agents in the process of globalization. Human agents (individuals, or collections of individuals organized in groups, governments, and states) appear as mere bystanders; and can act only within the limits set by markets and the 'logic' of capital.

However, a third position has emerged that represents one that is intermediate between individualist and holist positions. It insists that explanations of 'macro' social phenomena must be supplied with 'micro-foundations'—with an account of the intentional states that motivate individual action (Little 2000). This involves combining individualist and holist ontologies and methodologies through developing arguments that identify macro–micro linkages. These linkages help us, not only to explain macro-social phenomena, but to explain, as well, how macro-social phenomena produce social outcomes.

3. How do macro-social 'social mechanisms' produce social outcomes? Providing explanations of macro-social phenomena with micro-foundations

How do we provide explanations of macro-social phenomena (for example, capitalism or democracy)—with micro-foundations—with an account of the role of individual actors in producing it? An influential answer to this question is one offered by James Coleman. Coleman introduced a diagram, often referred to as 'Coleman's Bathtub' or 'boat' (because of its trapezoidal shape), that depicts a series of macro-micro linkages which, together, provide an explanation, both of the role of human agency in producing macro-social phenomena, and of how macro-social mechanisms produce social outcomes. What the diagram depicts is that (1) causal relations flowing downwards from *macro phenomena* (e.g. institutions) shape the *conditions of individual actions*; (2) these conditions then give rise to *individual actions*, and (3) individual actions in turn aggregate up to macro outcomes (Coleman 1990: 8).

To illustrate how, together these linkages provide what he calls a 'complete explanation', Coleman uses Max Weber's famous explanation, in *The Protestant Ethic and the Spirit of Capitalism* (1905), of how a macro-social phenomenon, Protestantism, contributed to bringing about another macro-level phenomenon: the rise of capitalist economic organization. In effect, Coleman was arguing that Weber's explanation of the rise of capitalism lacked 'micro-foundations'; that macro phenomena (e.g. capitalism) have to be explained, not by some other macro-level phenomenon (Protestantism), but by *the interaction of micro and macro levels*.

Figure 2.3 presents a schematic representation of Weber's argument (and serves also as an illustration of methodological holism). Coleman argued that social change comes about through changes in individuals and their interactions, so explanations of social change that refer *only to macro-level factors* are incomplete. Figure 2.4 depicts the micro-level changes that, if combined with Weber's macro-level variables, might provide a complete explanation of the rise of capitalism (and it illustrates, as well, methodological individualism).

Figure 2.3 Methodological holism

Source: Reproduced with permission from Coleman, J.S. Social Theory, Social Research, and a Theory of Action. *The American Journal of Sociology*, 91(6): 1309–1335. https://doi.org/10.1086/228423. Copyright © 1986, The University of Chicago Press Journals.

Figure 2.4 Methodological individualism

Source: Reproduced with permission from Coleman, J.S. Social Theory, Social Research, and a Theory of Action. *The American Journal of Sociology*, 91(6): 1309–1335. https://doi.org/10.1086/228423. Copyright © 1986, The University of Chicago Press Journals.

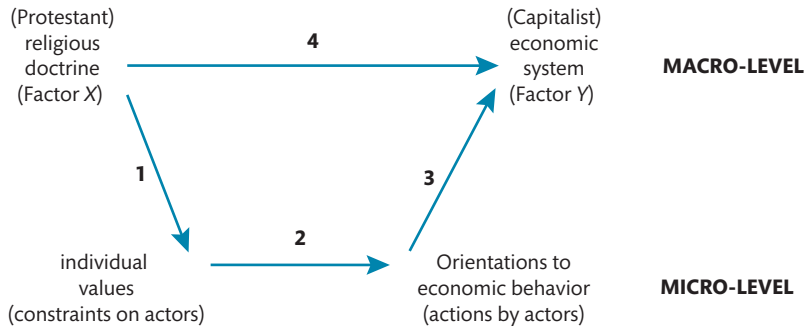


Figure 2.5 Coleman's Bathtub: a depiction of macro–micro linkages

Source: Reproduced with permission from Coleman, J.S. Social Theory, Social Research, and a Theory of Action. *The American Journal of Sociology*, 91(6): 1309–1335. https://doi.org/10.1086/228423. Copyright © 1986, The University of Chicago Press Journals.

Coleman then suggests how, by combining macro-level and micro-level variables and thus providing Weber's argument with 'micro-foundations', a 'complete explanation' of the rise of capitalism might be constructed. This is shown in Figure 2.5, where Coleman uses his 'bathtub' to suggest linkages among the macro-level and micro-level variables shown in Figures 2.3 and 2.4. and how they combine to explain the rise of capitalism. Weber explained Macro Factor *Y* (e.g. capitalism) in terms of Macro Factor *X* (Protestant religious doctrine). But as Coleman argued, this explanation is incomplete: a full explanation must explain how Protestant religious doctrine affected the values of individuals (a transition from the macro to the micro level), how these affected the actions and interactions of individuals (at the micro level) and how, in turn, these actions contributed to the rise of capitalism (a transition from the micro to the macro level). In sum, a full explanation requires an explanation of the macro-to-micro and the micro-to-macro transitions: how Macro Factor *X* creates constraints on actors (arrow 1); how actors choose actions under the constraints (arrow 2); and how the actions aggregate to the macro level (arrow 3). Using Coleman's scheme, Peter Hedström and Richard Swedberg provide a useful summation of the process through which social mechanisms produce outcomes through macro–micro and micro–macro interactions. The process begins by identifying a macro phenomenon that has micro implications, followed by a causal process on the micro level that leads to a new micro level, followed finally by a macro-social change that reflects the micro-social situation (Hedström and Swedberg 1998).

We have been discussing the question of the relationship between agents (the actors) and the structures which shape, give meaning to, or make possible their actions. It is worth noting that this question is at the core of what is known as the 'structure–agency problem' in International Relations. The 'problem', however, is one that arises in all social research. All research must confront the problem of defining how human agents act in social life, the social structure or system within which they act, and the relationship between the two (see Box 2.7). While we cannot offer further discussion of this issue here, we will return to it further on in this chapter. But for a full discussion of the debate and its implications, please visit our Online Resources.

We now have introduced some of the key terms of reference in a continuing debate about the scientific status of unobservable elements of social life. But, before moving on from discussion of scientific realism to a consideration of other non-positivist approaches, it is worth noting a related position that has emerged in the field of politics: **critical realism**.

BOX 2.7 The Structure-Agency Debate

1 Ontological question	Which has prior or primary ontological status (which came first): agents or structures? Do social structures exist independently from individuals?
2 Methodological question	What is the proper focus of social scientific explanation? Should explanations of social phenomena be expressed in terms of individuals and relations between them; or can they also invoke social phenomena (social structures, institutions, or cultures)?

For further discussion of these issues, see the Online Resources.

Critical realists share with scientific realism the assumption that, in addition to observable elements of life, reality also consists of unobservable elements, and that we know they are 'real' because they generate observable outcomes. As we have seen, it is this conception of realism that, not only distinguishes scientific realism from positivism: it is what makes scientific realism a *non*-positivist approach. For critical realists, this non-positivist conception of realism provides a basis for approaches to social inquiry that are not only non-positivist but also *critical*; for, in their view, exploring unobservable structures that impact social life opens up the possibility of our being able to change our world (Bhaskar 1998: 2). However, it rejects a key assumption that scientific realism *shares* with positivism: that there exists a reality separate from our perception or observation of it. Realist approaches that make this assumption represent what critical realists call 'naïve realism'. They argue that, since perception is a function of the human mind, we can only acquire knowledge of the external world through critical reflection on perception: that is, the way we become aware of, understand, or interpret something. It seems to us that, while some political researchers see the terms 'scientific realism' and 'critical realism' as synonymous (see, e.g., Brown 2007: 409), the emphasis that critical realism places on perception moves us further in the direction of the interpretivist approaches that we will be discussing next. In Chapter 14, we provide some examples of how a critical realist position informs political research in our discussion of **critical discourse analysis**—a type of textual analysis inspired by, and to a large degree consistent with, a critical realist philosophy of science.

We can conclude our discussion of scientific realism here. We have said that, as an approach to social inquiry, it provides an alternative to positivism; and that it differs from positivism in that, rather than relying solely on direct observation to acquire knowledge of the world, it assumes that reality consists of both observable and unobservable elements. But we have also shown that, in some respects, scientific realism and positivism are more similar than dissimilar. Most importantly, they both assume that the world exists independently of our knowledge of it. As we shall see in the next section, this is an assumption that has important implications for what we treat as legitimate knowledge and how we conduct research. We turn now in the next section, to a set of approaches, commonly grouped under the term 'intepretivism', which breaks decisively with this assumption and which promotes ontological and epistemological positions that stand in diametric opposition to those of positivism. Before turning to this discussion, it may be useful to review the various positions just discussed, and situate positivism within them.

Question: Can the methods developed by scientists to gain knowledge of the natural world enable us to produce reliable and objective knowledge of the social world?

Answers	Philosophical Position	Elaborations of this Positionn
Yes	**Positivism**: There are no fundamental differences between the natural and the social worlds, and so there should be none between the natural and social sciences (naturalism). What we know of the world is limited to what can be observed (empiricism), and laws govern the social world just as they govern the natural world; so social phenomena can therefore be explained through the discovery of laws.	**Logical positivism** Adds the assumption that we gain knowledge of the social world, not only through what can be observed (empiricism), but also through logical reasoning.
	***	***
	Scientific realism: Reality consists, not only of observable elements of life, it also consists of unobservable elements, and we know these unobservable elements are 'real' because they generate observable outcomes. There are knowable, mind-independent facts, objects, or properties that cannot be directly observed but which are, nonetheless, *real*. Unobservable elements of social life, such as structural relations between social phenomena, are crucial to an understanding and explanation of what goes on in the world. This is precisely what makes the social world similar to the natural world.	**Critical realism**: Rejects the assumption, of both positivism and scientific realism, that there exists a 'mind-independent' reality that is separate from our perception or observation of it. Perception is a function of the human mind, so we can only acquire knowledge of the external world through critical reflection on the way we become aware of, understand, or interpret something.
No	**Interpretivism:** The social world is fundamentally different from the world of natural phenomena because social phenomena are *subjectively* created: they are what we experience them to be and, thus, are socially or discursively constructed. Consequently, the primary goal of social science must be to achieve an understanding of human behaviour through an interpretation of the meanings, beliefs, and ideas that give people reasons for acting.	

Interpretivism

Interpretivism maintains that the social world is fundamentally different from the world of natural phenomena and that we, therefore, cannot gain knowledge of it by employing the methods used to explain the natural world. We cannot understand the social world by searching for *objective* regularities of behaviour that can be summed up in social scientific laws analogous to the laws of physics, because it does not exist independently of our interpretation of it. Social phenomena are *subjectively* created: they are what we experience them to be and, thus, are socially or discursively constructed. Consequently, rather than seeking to explain and predict social phenomena by means of laws, the primary goal of social science must be to achieve an understanding of human behaviour through an interpretation of the meanings, beliefs, and ideas that give people reasons for acting. To this requires that we employ a hermeneutical approach.

Hermeneutical approaches in the social sciences assume that human behaviour is the product of the meanings and intentions actors employ. Their goal, therefore, is to interpret social phenomena by developing an understanding of the meanings that actors give to their actions. Interpretivists argue that, to study human behaviour it is necessary, not only to employ a hermeneutical approach, but also to reject empiricist scientific methods. Empiricist methods treat social reality as consisting only of what Taylor calls 'brute data'. By 'brute data', Taylor means 'data whose validity cannot be questioned by offering another interpretation or reading, data whose credibility cannot be founded or undermined by further reasoning' (1994: 184). These data capture political behaviour involving actions that have an identifiable physical end state. When actors raise their hands at a meeting at the appropriate time we can treat this action as 'brute data' and say that the actors are 'voting for the motion'. However, the action may have meanings for the actors that are not captured by this description. When an actor votes for a motion, they may also be expressing loyalty to their party or defending the value of free speech (Taylor 1994: 190). But a 'behavioural' (or positive, or empiricist) political science deals only with brute data and their logical consequences, and not with the meaning of political behaviour for the actors themselves.

Let's consider the implications of this view for how we conduct political research.

Recall our earlier discussion about behaviouralism. Behaviouralist research is positivist and empiricist. Its concern is with what political actors do and why they do it. It seeks to discover the causes of behavioural outcomes by understanding the motivations of political actors. It uses public opinion polls and social surveys to learn about the beliefs, attitudes, and values that motivate behaviour, or rational choice theory to explain how individual self-interested rational action motivates behaviour. However if, as interpretivists contend, people act on the basis of the meanings they attach to their own and to others' actions, then understanding human behaviour requires an understanding of these meanings. Consequently, social science must be concerned, not with discovering causes of social outcomes, but with piecing together an interpretation of the meanings of a social outcome or production.

So, interpretivist approaches seek to understand human behaviour through forms of social science 'that emphasize understanding the meaning that social behaviour has for actors' (Gibbons 2006: 563). There are a multiplicity of interpretivist approaches used in the social sciences, most notably hermeneutics, cultural anthropology, *verstehen* social theory, critical theory, and **post-structuralism**. In what follows, we will highlight the differences between interpretivist and positivist approaches by focusing on hermeneutics.

'Hermeneutics' originally referred to a method used to interpret theological and legal texts (the German word 'hermeneutics' is translated into English as 'interpretation'). Today, 'hermeneutics' refers to theories and methods used to interpret texts of all kinds. These include not just written documents, but any object or practice that can be *treated as a text* and can, therefore, be the subject of interpretation. Interpretivists argue that human beings and their actions can be treated as a text and the subject of hermeneutical interpretation because, if behaviour is a product of the meanings and intentions employed by social actors, then the social scientist endeavouring to understand that behaviour is involved in an interpretive exercise not unlike that engaged in by the translator of a text.

The philosopher Charles Taylor elaborates this argument in an influential essay entitled 'Interpretation and the Sciences of Man' (1994). Taylor explains that any field of study can

be the object of hermeneutics if it meets two requirements: it must (1) contain an object or a field of objects that is a text, or a 'text-analogue', and (2) be in some way unclear (i.e. 'confused, incomplete, cloudy' or 'seemingly contradictory'; Taylor 1994: 181). Does the study of politics meet these criteria? Yes. We can treat the behaviour we are concerned to understand—the actions of a government, or the behaviour of members of a group towards one another—as 'texts'; and, since the motives and goals of this behaviour are often unclear or at odds with the pronouncements of the political actors involved, we can use interpretative methods to make sense of this behaviour.

As Taylor points out, brute data captures more than behaviour that has an identifiable end state: it also captures the subjective reality of individuals' beliefs, attitudes, and values 'as attested by their responses to certain forms of words or, in some cases, their overt non-verbal behaviour' (1994: 198–9). But while these data capture subjective meanings, there are *non*-subjective meanings constitutive of social reality that they cannot capture such as, for instance, intersubjective meanings and common meanings. Intersubjective meanings are meanings that do not exist only in the minds of agents but are rooted in and constitutive of social relations and practices such as paying taxes and voting. Common meanings involve recognition or consciousness of shared beliefs, aspirations, goals, values, and a common reference point for public life of a society (1994: 197). Taylor argues that we need to study these non-subjective meanings in order to comprehend political issues such as social cohesion, stability, disorder, and legitimacy. Moreover, they are crucial for 'a science of comparative politics': without them, we 'interpret all other societies in the categories of our own' (1994: 200), rendering invisible important differences between societies and making comparison impossible.

To stay with this point for a moment longer, analyses based on positivist epistemological assumptions, like those offered by rational choice theory, depend on an abstract description of human agency, one that pays little attention to differences across social, cultural, and historical settings. Rational choice theory seeks to explain social phenomena as the result of purposive choices by individuals within a given set of material and structural circumstances. As Daniel Little puts it: 'Agents like these in structures like those, produce outcomes like these' (2009). But interpretive approaches see individuals as unique, and human activities, actions, and social formations as unique historical expressions of human meaning and intention. Consequently, they are concerned, not with abstract descriptions of human agents, but with detailed interpretive work on specific cultures.

These differences between positivist and interpretive approaches can be illustrated by reference to a key area of research in our field: political participation. Positivist studies typically equate political participation with voting. However, interpretivists would argue that a particular voter may 'not understand voting as participation at all'. (It is neither practically nor logically impossible that an actor could say, 'No, I do not participate in politics, but I do vote' (Schwartz 1984: 1118).) But positivists tend to treat participatory acts as 'brute facts', as having an 'objective' ontological status: as existing 'in some sense "in the world" separate from the theoretical stance of the observer or of the participant' (Schwartz 1984: 1119).

Joel Schwartz points out that there is no objective point of view from which to describe and understand participation, that participation 'is a "subjective" phenomenon much like "justice" and "virtue"' (Schwartz 1984: 1119). While positivist studies typically equate political participation with voter turnout, participation involves a variety of political acts.

By imposing their own concept of 'participation', researchers are prevented 'from seeing the plural forms that participation in fact takes in the world. Whether acts (of an American voter or demonstrator, a French revolutionary, a Muslim revolutionary, a Solidarity member, and so on) count as acts of participation depends on those actors' subjective understanding of what they are doing' (Schwartz 1984: 1117).

In sum, interpretivism maintains that we cannot understand, explain, or predict patterns of human behaviour unless we understand the meanings and intentions that concrete agents attribute to their environment (social and natural); the values and goals they possess; the choices they perceive; and the way they interpret other individuals' social action. Humans act because of what things mean, so an understanding of human behaviour requires that we develop an understanding of the meanings they employ.

Clearly, there are important differences between the positivist and interpretivist approach to social inquiry. But the tendency to cast positivism and interpretivism approaches as 'two grand traditions in social science epistemology' (Pollins 2007: 93) exaggerates these differences. While they have different ontological and epistemological commitments, researchers working in both traditions generally follow the same methodological conventions, and so can understand what those working within the other tradition are doing. While researchers depend upon different assumptions, and may be interested in and test different questions, 'practical investigation of these questions often leads them to similar methodological tasks' (Finnemore and Sikkink 2001: 395). The methodological conventions they share include the following:

a. clear differentiation of premises from conclusions;
b. acknowledgement that sampling strategies matter;
c. recognition that some standards of validation must be established for the sources of evidence used;
d. differentiation of causes from correlations;
e. recognition that the spectre of spuriousness haunts all correlations;
f. acceptance of deductive logic;
g. belief in the need for the contestability of findings (Hopf 2007: 56).

These shared methodological conventions may, in fact, be seen as reflecting a common research practice founded in the hypothetico-deductive method (discussed previously). As Brian Pollins puts it: some researchers 'assess whether the information they have gathered fits with the interpretation they have posited', and others 'consider the fit of competing interpretations with the facts they have gathered', but 'in either case they are practicing the hypothetico-deductive method' (Pollins 2007: 100). In fact, according to Dagfinn Føllesdal, the hermeneutic method that we discussed in our consideration of interpretivism earlier 'is actually the hypothetico-deductive method' applied to materials that are 'meaningful': material that expresses an actor's beliefs and values (1994: 233). Interpretation-hypotheses can be judged by deducing consequences from them and confronting them with data, such as, for instance, a given text and related works bearing on it.

While interpretivists and positivists do not necessarily differ in their approaches to gathering relevant evidence, they do differ in their conception of what constitutes explanation

(recall our discussion, earlier, about political participation) and in their understanding of evidence.

To understand these differing conceptions of both explanation and evidence, it is useful to think about the difference defined between what might be called 'external' and 'internal' explanation and evidence. External explanations are associated with positivist research: they tend to work via correlations or deductions on the basis of *ascribed* reasons, and so need not concern themselves with actors' understandings of the world. Recall what we said about rational choice theory: it depends on an abstract description of human agency and so pays little attention to differences across social, cultural, and historical settings. In explaining social outcomes, it assumes that, in any particular set of circumstances or strategic interactions, the rational choices of individuals will produce predictable, law-like outcomes. Interpretive explanations, on the other hand, are 'internal' in the sense of their being concerned, not with abstract descriptions of human agents, but with the world of meanings inhabited by the actor (Hampsher-Monk and Hindmoor 2009: 48) and with detailed interpretive work on specific cultures. The distinction can be applied to different types of evidence, as well: 'external evidence' consists of empirical evidence about the behaviour, and the effects of the behaviour, of particular actors, while 'internal' or interpretive evidence consists of evidence about the beliefs of actors whose actions comprise the phenomena to be explained.

To further highlight these distinctions, below we compare an analysis based on rational choice theory (a positivist approach), constructivism (an interpretivist approach), and critical realism with respect to a particular area of political inquiry: the eruption of ethnic conflict in the former Yugoslavia.

The analysis of ethnic conflict: a positivist (rational choice), interpretivist (constructivist), and critical realist approach

The eruption of ethnic conflict in Yugoslavia and the subsequent collapse of the country, caught political obsevors by surprise. Among the many explanations offered by researchers in our field were a number of rational choice and constructivist explanations for why ethnic conflicts erupted in Yugoslavia in the 1990s. These explanations were concerned to challenge, and provide an alternative to, the cultural, 'ancient hatreds' explanation for the war between the Serbs and Croats. As many people have noted, the vast majority of Serbs and Croats had lived together peacefully during the previous sixty-year history of Yugoslavia. There is thus a key 'puzzle' that must be explained: why relations among communities that had been living together peacefully became polarized *so quickly*, before finally dissolving into savage violence.

Let's consider, first, an analysis which uses rational choice theory and is therefore based on positivist tenets. Rational choice theory explains outcomes as the result of rational choices made by individuals within a given set of material and structural circumstances. It shows that, given a particular set of circumstances, the strategic interactions of agents will produce predictable, law-like outcomes. Much of the analysis of strategic political interactions focuses on how individual actors make decisions in game-like situations. A 'game' is any situation in which a fixed set of agents or players, with a fixed set of strategies available to them, compete against one another, and receive a payoff as a result of the strategies

they and their fellow actors choose. The assumption is that all players know all possible outcomes and have preferences regarding these possible outcomes based on the amount of value or utility they derive from each of them. All players behave rationally; and they make rational decisions about what strategy to pursue based on a calculation of the costs and benefits of different strategies for achieving their preferred outcome.

As James Fearon points out, just months before the eruption of war between Serbs and Croats in Croatia, journalists had reported that 'most people seemed to have had no use for, or interest in, the exclusivist arguments pushed by the minority of extremists' (1998: 114). However, in spring 1991, 'Serbs and Croats who had resisted the extremists' appeals finally opted for division and war' (Fearon 1998: 115). To explain the 'puzzle' of why Serbs and Croats in Croatia became rapidly polarized in 1991, Fearon developed a game-theoretical model (a model based on assumptions of game theory) of how groups of people interact when confronted with a situation of uncertainty. Fearon argued that ethnic conflict erupted in Yugoslavia because, with the collapse of the Soviet Union, there emerged a 'commitment problem'. A 'commitment problem' arises when 'two political communities find themselves without a third party (e.g. a strong state or powerful ally) that can guarantee agreements between them' (1995: 2). If ethnic minorities don't believe that the state can guarantee that the ethnic majority will not infringe on their rights, they will prefer to fight for succession while the state is still weak. This, he argues, helps to explain the rapid polarization of ethnic groups in Croatia following the declaration of Croatian independence in June 1991.

Note how this explanation is consistent with the hypothetico-deductive model. It consists of (1) a set of initial determining conditions (circumstances pertaining at particular times and places); and (2) a general law or laws which connect these conditions to the type of events to be explained ('hypotheses' which are capable of being confirmed or disconfirmed by suitable empirical findings). First, Fearon specifies the set of initial conditions in which the hypothesized commitment problem might be expected to arise. It arises, Fearon tells us, whenever three conditions hold: (1) the groups interact in an environment where no third party can guarantee and enforce agreements between them; (2) one of the groups anticipates that its ability to secede or withdraw from joint arrangements will decline in the near future; and (3) for this group, fighting in the present is preferable to the worst political outcome it could face if it chooses continued interaction (1995: 10). Second, and on the basis of his game-theoretic model, Fearon generates hypotheses about what, with the emergence of a commitment problem, makes ethnic war more or less likely. The key mechanisms include (1) change in the size of the relative military power between groups when a new state is formed; (2) the relative size of the ethnic minority; (3) whether majority and minority groups' costs for fighting are low; and (4) whether institutions can be created that give minority groups political power that is at least proportional to their numbers.

Fearon then applies the model to the war in Croatia in 1991–2. What he describes as a 'commitment problem' arises for Serbs when Croatia declared its independence from Yugoslavia and the Serbian minority living in Croatia were facing the prospect of being in a state with no credible guarantees of their political status, or their economic and even physical security. What evidence does Fearon offer for the emergence of a commitment problem in Croatia? Fearon states that if a commitment problem was involved in what was going on in Croatia, we would expect to find that Croatian leaders were trying to work out guarantees with Serb leaders' (1998: 119). This, in fact, is what Fearon finds occurred on numerous

occasions. However, the efforts of Croatian President Tudjman to solve the commitment problem were unsuccessful. Consequently for Serbs, facing the 'prospect of entering the new state of Croatia with no credible guarantees of their political status or security', resorting to war appeared a better prospect than the prospect of fighting later, by which time the Croatian state would have grown stronger (Fearon 1998: 116).

Fearon's evidence, then, consists in showing that there is a 'fit' between the deductions of the theory and the observed behavioural outcomes—that the outcome is consistent with the theoretical predictions. Fearon also endeavours to demonstrate that his explanation offers a better 'fit' with the facts than other explanations do. Finally, he suggests that his findings in this case are generalizable to other cases. The basic commitment problem that he describes, he states, 'appears either to lurk or to have caused interethnic violence' in other cases, as well: in Azerbaijan, Georgia, Moldova, Ukraine, Estonia, Zimbabwe, South Africa, and Northern Ireland (Fearon 1995: 21).

Let's consider how an interpretivist approach, **constructivism**, explains the rapid ethnic polarization that occurred in the former Yugoslavia in the 1990s.

Constructivism is an approach that has had an important influence on political inquiry. Consistent with interpretivist assumptions, constructivism maintains that reality does not exist as something independent of us and is not, therefore, merely discovered by us: it is socially, and actively, *constructed*. Social phenomena are social constructs, constructivists assume, in the sense that their shape and form is imbued with social values, norms, and assumptions, rather than being the product of purely individual thought or meaning. Actors are not totally free to choose their circumstances. They make choices in the process of interacting with others and, as a result, bring historically, culturally, and politically distinct 'realities' into being. States and other social institutions also take specific historical, cultural, and political forms that are a product of human interaction in the social world. In this respect, the world of politics is a social construction rather than something that exists independently of human meaning and action.

In contrast with positivist approaches which emphasize a single objective reality, the idea of social construction suggests differences across contexts. It is not only the material environment, but also the cultural and institutional environment that provides incentives and disincentives for behaviour. Society is more than just the site in which actors pursue predefined interests in a rational, utility-maximizing manner. Rather, it is an environment that forms and influences the identities and interests of actors and makes them who they are. The identities and interests of actors are not intrinsic to them; they are socially contingent, they depend on social interaction, bargaining/negotiating, arguing, communicating in general. Both the identities and interests of actors are constituted (formed, influenced) through interaction and by the institutionalized norms, values, and ideas of society. Since the interests and identities of actors are not given—but result from social interaction—they cannot be abstracted from the social conditions which produce them; and they are subject to change as a result of political processes.

Like rational choice approaches, constructivism rejects explanations of nationalist and ethnic phenomena as the outcome of essential cultural identities. But, unlike rational choice approaches, it emphasizes that identities and interests, and the role of identities are multiple and fluid and politically malleable—influenced by surrounding structures and 'constructed' for political purposes.

Let's consider how Murat Somer (2001) uses this interpretive approach to address the same 'puzzle' as Fearon: the rapid ethnic polarization that occurred in Yugoslavia in the 1990s. Somer emphasizes the importance that public discourses play in forming individuals' ethnic identities, and in suppressing and reviving dominant perceptions of ethnic identities. He argues that ethnic conflict arises when processes of ethnic identity construction in the public arena construct a divisive image of identities. Ethnic polarization in public discourses is immediately observed and immediately affects the decisions of groups within society.

The dominant image of ethnic categories in society changes as a result of cascades of individual reactions; and once this cascading process begins to change behaviour and attitudes it is very difficult to stop. In Yugoslavia, ethnic polarization in public discourses was engineered by ethnic entrepreneurs who constructed and promoted a divisive image of ethnic identities as mutually exclusive and incompatible with belonging to the same nation. This triggered a 'cascade process', which resulted in the creation of a critical mass of opinion around a new image of ethnic identities. People who had secretly held this divisive view, but also people who now felt compelled to support it, jumped on the bandwagon. Hence, the divisive image became the norm, and it became 'inappropriate, even blasphemous, to defend interethnic mixing and brotherhood' (Somer 2001: 128).

Somer highlights the role of public discourse in 'preference falsification'—in motivating people to publicly 'falsify' their private beliefs. During the communist era, state policies promoted a public discourse aimed at eradicating the public expression of a divisive image of ethnic relations in the country; but these policies did not succeed in eradicating privately-held divisive views of ethnic relations. Consequently, the *public* discourse promoted by the state in Yugoslavia had exerted pressure for 'downward falsification' by discouraging people from openly expressing their ethnic prejudices. This downward preference falsification concealed, to most observers, the *private* importance of the 'divisive image'; consequently, 'even analysts who had a fair idea about the private significance of the divisive image were surprised by the severity of polarization' in the 1990s (2001: 136). During the 1990s, 'the dominant public discourse emphasizing unity and brotherhood' turned into one that emphasized 'radical ethnonationalism' (2001: 136). But this 'public polarization far exceeded private polarization' (2001: 143). Consequently, there was widespread 'upward ethnic preference falsification', the exaggeration of public support for the divisive image, as the new nationalist regime exerted pressure for people—including liberal and moderately tolerant individuals—to think and act in an ethnically intolerant manner.

Where did Somer look for and find evidence of this? Somer uses survey research which indicates decreases in self-identification with the overarching Yugoslav identity between 1981 and 1991. The respondents were anonymous, so these surveys were able to capture changes in people's private preferences. During the 1980s there was a striking upsurge in 'the public expression of the divisive image' (2001: 143). But, 'in 1989, when public polarization had reached an advanced state, anonymous surveys continued to reveal that interethnic tolerance levels were high by global standards'. So, 'while the public discourse was becoming increasingly more divisive and less tolerant' of ethnic

differences, 'private attitudes remained quite tolerant' of them. In fact, 'the highest levels of tolerance were found in Bosnia, the site of the most violent crimes'. This tolerance is shown in the large number of people who deserted and evaded the military call-up. In fact, '[d]esertion and call-up evasion were very common during the civil war when public support for the divisive image was at its peak' (2001: 144).

So, we have looked at two approaches to understanding ethnic polarization in Yugoslavia. We might say that both are constructivist in the sense that they see changes in actors' identities as likely to occur in moments of crisis and dilemma. But Fearon, consistent with rational choice approaches, emphasizes the role of strategic calculation in identity construction, while Somer emphasizes cognitive features, such as norms of behaviour and intersubjective understandings (though these don't necessarily operate to the exclusion of the calculative element stressed in the rational choice explanations offered by Fearon and others).

In comparing the two analyses we can see that they provide an illustration of the distinction between 'internal' and 'external' explanation and evidence that we previously discussed.

Fearon's analysis draws on the assumptions of rational choice theory to model external 'behaviour' and then seeks evidence by way of deductive fit with that model. Empirical evidence consists of statements and activities of Croatian leaders that indicate a concern for providing the Serb minority with a commitment to guarantee their status and cultural autonomy. But there is no direct evidence of the existence of a commitment problem among the Serb population, of strategic behaviour on the part of individuals or groups, or of the relationship between belief and action. Direct evidence concerning whether the individual or group choices which led to ethnic polarization were made for the reasons stated in his model might be impractical or impossible to obtain. Instead, the 'test' of the model involves (1) deducing the factors that, in the given circumstances, might be expected to lead groups to resort to violence; and (2) observing the outcomes. Where we observe that circumstances favour the behaviours that the model tells us are most likely to occur, we can infer that there is a line of cause and effect that relates those circumstances and the outcome.

Somer's analysis draws on the assumptions of constructivism. In contrast to rational choice approaches, which assume that agents act on the basis of fixed interests and preferences, constructivism assumes that identities and interests are a product of human interaction. Consistent with this assumption, Somer seeks, through exploring 'cascade' processes, to understand how public political discourses form and influence the identities and interests of actors. He then combines survey data with examination of the historical events to provide evidence of the changing nature of public and private views and to suggest the relationship between them.

Let's now consider a critical realist analysis of the ethnic conflict that erupted in the former Yugoslavia. Like scientific realism, critical realism maintains that explanation can be based both on observable elements of social life as well as unobservable social structures or mechanisms. However, unlike scientific realism, critical realism insists that these and other social phenomena 'do not exist independently of the agents' conceptions of what they are doing in their activity' (Bhaskar 1979: 48).

As we previously discussed, the idea that unobservable 'social mechanisms' produce outcomes in the social world renewed debates about what we identified as the 'structure–agency problem': the problem of defining (1) how human agents act in social life, (2) the social structure or system within which they act, and (3) the relationship between the two. Critical realism assumes that structures are the ever-present condition of human agency, and that *agents* interact with, and thereby reproduce or transform the *structures* that enable and constrain their actions. An influential conception of how objective structure and individual, subjective agency interact over time is Margaret Archer's (1995) 'morphological model' in which *structural conditioning* (which is treated as temporally prior) conditions *social interaction*, which in turn generates *structural elaboration*, as shown below.

(T= time; T1 refers to some prior/initial time; while T2, T3, T4 refer to later points in time).

Structural conditioning
*T1

Social interaction
T2 T3

Structural elaboration
T4

The Morphological Approach: how structures and agents impact each other

The two explanations of the Yugoslav conflict we've discussed both tend to emphasis the role and significance of agents and, in particular, powerful leaders: Fearon emphasizes that it is *political leaders* (Serb and Croatian) who 'foment … interethnic hostility' (1998: 113); Somer argues that 'divisive *ethnic entrepreneurs*' were decisive in ethnic polarization (2001: 130) by promoting a 'public discourse' that became 'increasingly more divisive and intolerant of interethnic differences' (2001: 144). But neither analysis explores the embeddedness of these powerful elites in antecedent structures.

A critical realist analysis of these events would start with the elite discourses which Fearon and Somer emphasize, but would engage with more probing questions about them. Because it is concerned with the structures that 'pre-exist agents and provide the conditions for human action', and how agents interact with structures to reproduce or transform them and themselves as agents, a critical realist reasoning would force the analysis to 'move backwards', to think about the conditions that are necessarily connected to these discourses, and without which they would not have existed (Fiaz 2014: 503). This 'moving backwards' is called 'retroduction', an inferential process in which the analysis moves backward from some initial puzzling fact/social interaction to the mechanisms which produced it. We move 'backwards' by asking 'what antecedent conditions/structures would make this phenomenon possible? What would account for/make comprehensible the phenomenon [social interaction] concerned?'[2]

This inquiry would lead us to explore a set of hypotheses about how some antecedent conditions or structures combined to produce the observed behaviour. As the accounts of Fearon and Somer suggest, these discourses emerged as the result of the purposeful actions of powerful political actors. So we might, for instance, draw on arguments advanced by V. P. Gagnon to hypothesize the following: that (1) 'violent conflict along ethnic cleavages is provoked by elites in order to create a domestic political context where ethnicity is the only politically relevant identity'; that (2) '[s]uch a strategy is a response by ruling elites to shifts in the structure of domestic political and economic power'; and that (3) by constructing 'individual interest in terms of the threat to the group, endangered elites can fend off domestic challengers who seek to mobilize the population against the status quo, and can better position themselves to deal with future challenges' (Gagnon 1994–5: 132). We might then investigate this argument through an in-depth exploration of Yugoslavia's social structures, the place of ethnic leaders in these structures, and how changes in them impacted these leaders.

A critical realist approach to explaining the Yugoslav conflict

The structures which provided the conditions for the actions of these elites, included (1) the emergence of 'secondary' economies within Yugoslavia, and (2) the republicanizarion of banks and political interests by the constitution of 1974. Yugoslavia's 'planned economy' was, 'in reality' a system of 'low intensity chaos': its five-year plans worked only if 'everyone reached their targets simultaneously'; and since 'delays and holdups were bound to happen, the system was inevitably in a state of perpetual breakdown' (Palan 1991: 122). 'Leading politburo and party members' used their political power and contacts to establish a secondary 'personalized' system that ensured 'the flow of credit, raw materials, spare parts, and manufactured goods' (1991: 122). Constitutional reforms introduced in 1974, shifted power from the central government to the republics, each of which was given a president, prime minister, government, and parliament. Although Yugoslavia remained formally a federal state, 'the basis was established for what [was] in practice a confederate structure as the six republics and the two autonomous regions were each granted what amounts to a separate central bank' (Palan 1991: 120). Together, the power relations that developed around the secondary economies, local governing elites, and local banking interests set the stage for the social interactions that developed following the death of Tito.

When Tito died in 1980, the conditions were in place for the '"feudalization" of Yugoslavia on (con)federate lines' (Palan 1991: 120). The secondary system offered abundant opportunities for corruption and, in post-Tito Yugoslavia, became thoroughly corrupt. An alliance of the upper echelon of the banks, industry, and the communist party 'gradually began to treat state industries as sources for private accumulation' (Palan 1991: 120–1), turning industries and services into 'virtual fiefs' of leading politicians. The party could 'carry on corrupt practices because there was no political mechanism to replace them; directors of industry and services were able to cream off profits because they were responsible to corrupt politicians; banks were providing credit to unviable industries because they were in effect members of a fief' (Palan 1991: 122).

> Fiscal and monetary policies were supposed to be run by the federal government. But the presidency was paralyzed by a lack of formal or informal agreement among the republics, and the economy was dragged down as money was syphoned off in an array of corrupt or ineffective investments.
>
> (Palan 1991: 121)

By 1989 the federal government under the premiership of the Croatian Ante Markoviç was itself championing the cause of reform. Markoviç visited Washington and obtained a solemn promise from President Bush to put the Yugoslavian banking system on a sound footing (Palan 1991: 121). 'A sound financial system, however, could short-circuit the whole process … and ultimately remove the basis of the politicians' power'. Markoviç's plan was to choke off the system 'at its most vulnerable point: the banking system'. The realization of this plan would have resulted in a centralized federal state, and 'the local "mafias" had no intention of giving up power without a fight' (Palan 1991: 123). The structural conditions which conditioned the social interaction through which those structures were both transformed and reproduced, is shown below:

T1: Structural Conditioning		T2-T3: Social Interaction				T4: Structural Elaboration
Emergence of a 'secondary economy', offering elites private access to resources and opportunities for corruption; republicanization of banks and political interests in the new 1974 constitution; weakening of centralized state power.	→	Government seeks to recentralize state power, threatening the position of powerful local elites.	→	Elites seek to alter public perceptions in order to counter this threat.	→	Eruption of ethnic conflict and collapse of the country.	→	Transformation of state structures and reproduction of local power structures.

Which of these different explanations for ethnic conflict we choose to favour has implications for politics and policy—for how people relate to each other and how governments act (Box 2.8). The choice of which set of assumptions provide the best starting point for your own research on a specific political issue or problem is therefore one which you will need to carefully consider.

BOX 2.8 Understanding Ethnic Polarization in Yugoslavia: Three Approaches

	Positivist (rational choice) approach	Interpretivist (constructivist) approach	Criticial realist approach
What determines an individual's identity?	Structural and constraining features of institutions	Social and cognitive aspects of actors responses to an institutional setting	Interactions with other individuals, many of which are determined by enduring social structures and institutional arrangements
How are these identities formed?	Through interaction and strategic calculation	Through norms of behaviour and intersubjective understandings, as well as strategic calculation	Through an an interplay between people, groups, organizations, political and economic systems, and social structures
What interests do agents pursue?	Agents act on the basis of fixed interests and preferences	Agents act on the basis of interests which emerge from identities that develop from the image they hold of themselves and of others	Interests take shape through agents' concrete social development that is conducted through stable social arrangements and institutions
Nature of explanation and type of evidence	External 'behaviour' is modelled and then evidence is sought in the fit between observed behaviour and the expectations of that model	A mechanism is identified linking popular beliefs, public discourse, and relations across groups, then evidence is sought from survey data as well as from examination of the historical events	Identification of the causal mechanisms/ broader structural determinants that combined to produce a given phenomenon, then testing the hypothesized mechanisms against empirical evidence from an in-depth case study, interviews, and other methods.

Conclusions

This chapter has begun our consideration of some fundamental ontological, epistemological, and methodological problems posed by social inquiry. These problems mostly branch out from one central question: are the methods of the social sciences essentially the same as, or essentially different from, those of the natural sciences? We have reviewed three different

and important answers to this question: positivism, scientific realism, and interpretivism. How these answers differ is presented in Box 2.9.

Each of these defines a particular position with respect to how we conduct research in politics. As Box 2.9 shows, each is based on fundamentally different assumptions about how we can know and develop knowledge about the social world. All of them remain important perspectives for contemporary social research. The differing ways in which they answer the question of whether and how we can pursue a systematic and rigorous study of social phenomena in the way that scientists pursue study of the natural world, and address philosophical presuppositions about 'reality' implicit in social science research, bear on how we design or structure an approach to research into political phenomena, and the claims we can make with respect to the findings that result from that research.

BOX 2.9 Positivism, Scientific Realism, and Interpretivism Compared

	Positivism	Scientific realism	Interpretivism
Ontology: What is the nature of the social world?	No different from the natural world: an objective reality exists independently of our knowledge of it (a 'naturalist' ontology).	No different from the natural world: an objective reality that exists independently of our knowledge of it (a 'naturalist' ontology).	Fundamentally different from the natural world. Reality is not mind-independent: it is subjectively created.
Epistemology: What sort of knowledge of the social world is possible?	Scientific knowledge of the social world is limited to what can be observed. We can explain and predict social phenomena by discovering empirical regularities, formulating law-like generalizations, and establishing causal relationships.	Scientific knowledge is not limited to what can be observed but also includes theoretical entities (unobservable elements of social life). We can explain and predict social phenomena based on theories about these entities and certify the truth of these theories by employing various rules of method.	Scientific knowledge can be gained through interpreting the meanings which give people reasons for acting. We can, in this way, understand human behaviour; but we cannot explain or predict it on the basis of law-like generalizations and establishing the existence of causal relationships.
Causality: What do we mean by 'causes'?	A reflection of the way we think about the world; established by discovering observable regularities.	A reflection of reality; established by discovering unobservable underlying generative mechanisms.	We cannot seek causes, but only uncover meanings that provide the reasons for action.
Methodology: How can we gain knowledge of the social world?	Through direct observation.	Through direct observation and logic applied to both observable and unobservable structures.	Interpretive theory and textual strategies. The social world is like a text and has to be interpreted to discover hidden meanings and sub-texts.

At the heart of the debate among these perspectives is the question of what sort of knowledge we can gain about social phenomena. This question is also central to the controversy that we take up in Chapter 3, which is the debate about whether knowledge produced through the study of the social world is or can be 'objective'.

Reality is constantly being defined for us—by political scientists, by historians, by politicians in their speeches, by media analysts in their news reports. The ability to identify the underlying structure of assumptions or the implicit theory which shapes a given account of reality, whether presented by scholars, politicians, or journalists, allows us to become more active analysts of contemporary politics, rather than depending on the analysis of others.

Questions

1. What is involved in providing an explanation of social phenomena? How is explanation distinct from and related to interpretation?
2. There are different conceptions of causality. What are these different conceptions? How, according to different conceptions of causality, do we establish that something causes something else?
3. How is describing, interpreting, and explaining human action different from describing, interpreting, and explaining non-human events?
4. Should the primary goal of social science be to provide law-like explanations, capable of supporting predictions? Or are law-like explanations impossible, or unnecessary, in social science?

Guide to Further Reading

Gibbons, M. T. (2006), 'Hermeneutics, Political Inquiry, and Practical Reason: An Evolving Challenge to Political Science', *American Political Science Review* 100(4) (November): 563–71.
This reviews the evolution of interpretivist political inquiry and how adherents of dominant approaches have responded to it, and it offers a valuable discussion of positivist and interpretivist assumptions and their implications for political research.

March, D. and P. Furlong (2002), 'A Skin Not a Sweater: Ontology and Epistemology in Political Science', in D. Marsh and G. Stoker (eds), *Theory and Methods in Political Science*, 2nd edition (Basingstoke: Palgrave), 17–41.
The authors discuss positivism, interpretivism, constructivism, and also realism as different approaches to ontology and epistemology in political science, and illustrate their differences with case studies.

Martin, M. and L. C. McIntyre (eds) (1994), *Readings in the Philosophy of Social Science* (New York: MIT Press).
This volume brings together a collection of important texts on the disputed role of general laws in social scientific explanation (Part II), and on interpretation and meaning (Part III).

Lane, R. (1996), 'Positivism, Scientific Realism and Political Science', *Journal of Theoretical Politics* 8(3): 361–82.
This article explores the implications of scientific realist principles for political science, political research, and political theory, providing examples of a scientific realist approach in studies utilizing a variety of theoretical approaches, including rational choice, new institutionalism, and comparative politics.

Little, D. (1991), *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science* (Boulder, CO: Westview Press), chapter 2, 13–38.

Russo, F. (2009), *Causality and Causal Modelling in the Social Sciences* (New York: Springer).
This book offers an overview of debates, and it provides a valuable analysis of reasoning about causation by looking at the causal arguments advanced in specific social science studies.

Sankey, H. (2008), *Scientific Realism and the Rationality of Science* (Aldershot: Ashgate), chapter 1.
This chapter provides a clear exposition of the doctrines of scientific realism, distinguishing between core and optional doctrines, and the principal arguments that have been advanced for scientific realism.

References

Amadae, S. M. and B. Bueno de Mesquita (1999), 'The Rochester School: The Origins of Positive Political Theory', *Annual Review of Political Science* 2: 269–95.

Archer, Margaret (1995), *Realist Social Theory: The Morphogenetic Approach* (Cambridge: Cambridge University Press).

Benoit, K. (2007), 'Electoral Laws as Political Consequences: Explaining the Origins and Change of Electoral Institutions', *Annual Review of Political Science* 10 (June): 363–90.

Bhaskar, Roy (1979), *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences* (Atlantic Highlands, NJ: Humanities Press).

Bhaskar, R. (1998), *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*, 3rd edition (Hemel Hempstead: Harvester Wheatsheaf).

Bhaskar, Roy ([1993] 2008), *Dialectic: The Pulse of Freedom*. London: Routledge.

Brown, C. (2007), 'Situating Critical Realism', *Millennium: Journal of International Studies* 35(2) (March): 409–16.

Coleman, J. S. (1986), 'Social Theory, Social Research, and a Theory of Action', *American Journal of Sociology* 91(6): 1309–35.

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

Colomer, J. P. (2005), 'It's Parties that Choose Electoral Systems (or, Duverger's Laws Upside Down)', *Political Studies* 53: 1–21.

Easton, D. (1969), 'The New Revolution in Political Science', *American Political Science Review* 63(4) (December): 1051–61.

Ekström, M. (1992), 'Causal Explanation of Social Action: The Contribution of Max Weber and of Critical Realism to a Generative View of Causal Explanation in Social Science', *Acta Sociologica: Journal of the Scandinavian Sociological Association* 35: 107–22.

Fearon, D. J. (1995), 'Rationalist Explanations for War', *International Organization* 49(3): 379–414.

Fearon, D. J. (1998), 'Commitment Problems and the Spread of Ethnic Conflict', in D. Lake and D. Rothchild (eds), *The International Spread of Ethnic Conflict* (Princeton, NJ: Princeton University Press), 107–26.

Fiaz, Nazya (2014), 'Constructivism Meets Critical Realism: Explaining Pakistan's State Practice in the Aftermath of 9/11', *European Journal of International Relations* 20(2): 491–515.

Finnemore, M. and K. Sikkink (2001), 'Taking Stock: The Constructivist Research Program in International Relations and Comparative Politics', *Annual Review of Political Science* 4: 391–416.

Føllesdal, D. (1994), 'Hermeneutics and the Hypothetico-Deductive Method', in M. Martin and L. C. McIntyre (eds), *Readings in the Philosophy of Social Science* (Cambridge, MA: MIT Press), 233–46.

Gagnon, V. P. (1994–5), 'Ethnic Nationalism and International Conflict: The Case of Serbia', *International Security* 19(3) (Winter): 130–66.

Gerring, J. (2007), 'Review Article: The Mechanismic Worldview. Thinking inside the Box', *British Journal of Political Science* 38: 161–79.

Gibbons, M. T. (2006), 'Hermeneutics, Political Inquiry, and Practical Reason: An Evolving Challenge to Political Science', *American Political Science Review* 100(4) (November): 563–71.

Hampsher-Monk, I. and A. Hindmoor (2009), 'Rational Choice and Interpretive Evidence: Caught between a Rock and a Hard Place?', *Political Studies* 58(1): 47–65.

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

Hempel, C. G. (1994), 'The Function of General Laws in History', in M. Martin and L. C. McIntyre (eds), *Readings in the Philosophy of Social Science* (Cambridge, MA: MIT Press), 43–54.

Hoffmann, M. J. (2006), 'Social (De) Construction: The Failure of a Multinational State', in J. A. Sterling-Folker (ed.), *Making Sense of International Relations Theory* (Boulder, CO: Lynne Reiner), 123–38.

Hopf, T. (2007), 'The Limits of Interpreting Evidence', in R. N. Lelbow and M. I. Lichbach (eds), *Theory and Evidence in Comparative Politics and International Relations* (Basingstoke: Palgrave/Macmillan), 55–84.
Hume, D. (1966), *Enquiries Concerning the Human Understanding* (Oxford: Clarendon Press).
Lane, R. (1996), 'Positivism, Scientific Realism and Political Science', *Journal of Theoretical Politics* 8(3): 361–82.
Leitenber Milton (1963), 'The Natural Birth of a Hypothesis', *American Behavioral Scientist* 7(3) (October): 3–9.
Little, D. (2000), 'Explaining Large-Scale Historical Change', *Philosophy of the Social Sciences* 30(1) (March): 89–112.
Little, D. (2009), 'McIntyre and Taylor on the Human Sciences', 16 July. http://understandingsociety.blogspot.com/2009/07/macintyre-and-taylor-on-human-sciences.html.
McIntyre, L. C. (1994), 'Complexity and Social Scientific Laws', in M. Martin and L. C. McIntyre (eds), *Readings in the Philosophy of Social Science* (Cambridge, MA: MIT Press), 131–44.
Palan, Ronen (1991), 'Misguided Nationalism: the Causes and Prospects for Slovenian Independence', *The Contemporary Review* 259(1058) (September): 119–27.
Pollins, B. (2007), 'Beyond Logical Positivism: Reframing King, Keohane, and Verba', in R. N. Lelbow and M. I. Lichbach (eds), *Theory and Evidence in Comparative Politics and International Relations* (Basingstoke: Palgrave/Macmillan), 87–106.
Riker, W. H. (1982), 'The Two-Party System and Duverger's Law: An Essay on the History of Political Science', *American Political Science Review* 76 (December): 753–66.
Ross, M. (2001), 'Does Oil Hinder Democracy?', *World Politics* 53: 325–61.
Schwartz, J. D. (1984), 'Participation and Multisubjective Understanding: An Interpretivist Approach to the Study of Political Participation', *Journal of Politics* 46: 1117–41.
Somer, M. (2001), 'Cascades of Ethnic Polarization: Lessons from Yugoslavia', *The ANNALS of the American Academy of Political and Social Science* 573(1): 127–51.
Taylor, C. (1994), 'Interpretation and the Sciences of Man', in M. Martin and L. C. McIntyre (eds), *Readings in the Philosophy of Social Science* (Cambridge, MA: MIT Press), 181–212.
Tilly, C. (2001), 'Mechanisms in Political Processes', *Annual Review of Political Science* 4: 21–41.
Trochim, W. M. (2006), *The Research Methods Knowledge Base*, 2nd edition. http://www.socialresearchmethods.net/kb/.

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

Endnotes

1. In recent years some researchers have modified Duverger's Law by suggesting that 'it is the number of parties that can explain the choice of electoral systems, rather than the other way round' (Colomer 2005: 1; see also Benoit 2007).
2. A critical realist analysis (1) describes a social phenomenon; (2) retroductively thinks of plausible mechanisms that explain why reality is this way; (3) eliminates rival theories by using judgemental rationality to decide which theory best explains the largest amount of the empirical evidence; and (4) 'tests' the efficacy of the theory, correcting the theory should contradictions arise through the process of testing it (Bhaskar [1993] 2008: 131). In this, the analyst can employ a range of evidence, including discourse analysis, interviewing, ethnography, case studies, comparative case methods, and historical methods.

3 Objectivity and Values

Chapter Summary

This chapter explores a key debate in the philosophy of social science: whether it is possible to separate facts and values in social science research. The debate raises fundamental questions about how values influence social-scientific inquiry, the role and responsibilities of the researcher in social-scientific inquiry, and the ways in which areas of study are shaped by the norms and values of research communities and environments. Consequently, the answers researchers give to these questions have important implications for how they approach research and understand its findings. Among the questions this chapter explores are the following:

- To what extent, and in what ways, do 'values' present problems for the analysis of the social world?
- Do values (either our own or those embedded in the disciplinary frameworks that we use) influence social inquiry?
- If so, in what ways, and to what extent, do values intrude into the research process?
- Do the research practices and methods of data collection and analysis that we employ ensure the objectivity of our findings?

Introduction

This chapter explores the question of whether social inquiry can be objective, or 'value-free'. This question is related to a number of other complicated and thought-provoking questions. Do values (either our own or those embedded in the disciplinary frameworks that we use) inevitably compromise the objectivity of social research? If so, in what ways and to what extent do they intrude into our research? Do, or can, the research practices and methods of data collection and analysis that we employ ensure that the findings we produce are objective? As we shall see, though different answers have been offered to these questions, they continue to be the focus of considerable debate. To be clear, the debate is not just about the particular set of research practices and methods that we assume enable us to produce objective research: it is about whether objective inquiry of the social world is even *possible*.

The discussion which follows explores this and a number of related questions. One of these questions is how, given that we choose to undertake research on issues that, in some way, are connected to our values or bear on matters that we regard as important, can we, in our research, avoid bias, misrepresentation, or the selective presentation of evidence? Is it possible for us to produce unbiased conclusions about the political issues we are investigating: to design and conduct research that is value-neutral, and to be scrupulously objective in the way in which we go about collecting and evaluating evidence?

There are many ways in which a researcher's biases can intrude on the research process. Being alert to the conscious and unconscious biases that we may have about a particular issue can prevent us from allowing our biases to influence our investigation of that issue. But there are broader, more complicated considerations that bear on the question of whether and how we can pursue value-free, objective research on social phenomena. One of these is the issue of whether it is possible to distinguish between or separate facts and values in scientific inquiry.

Chapter 2 discussed the different answers that positivism and interpretivism give to the question of whether it is possible to separate facts and values in social inquiry. Their answers differ because they make different ontological assumptions about the nature of what it is we seek to understand (the social world) and so are based on different philosophical presuppositions about 'reality'.

Positivism maintains that the social world is similar to the natural world in that, as in the natural world, there exists in the social world a mind-independent, objective reality that exists independently of our knowledge of it. It is therefore possible to employ the methods used to study the natural world in order to gain knowledge of the social world. Interpretivism maintains that the social world is fundamentally *different* from the natural world because reality in the social world *does not* exist independently of us but is in important respects subjectively created. Therefore knowledge of the social world *cannot* be gained through employing the methods used to study the natural world.

That facts are fundamentally different and distinguishable from values, and that they exist independently of the observer and the observer's values, is a central tenet of positivism. It is because we can distinguish between facts and values that science not only can be, but *is*, 'value-free'. Science is concerned with facts, not values; and considers that any intrusion of values in the research process contaminates the objective character of science. The distinction between facts and values is what differentiates science from non-science, and what defines the boundary between the 'scientific' study of society and ideology. Consequently, arguments that suggest it is not possible to distinguish between facts and values would seem to undermine the possibility of understanding the social world through scientific study—the possibility, that is, of social science.

There are a variety of theoretical approaches in our field that reject positivism, and the positivist assumption that a distinction can be drawn between facts and values. These approaches, often referred to collectively as '**Critical Theory**', include constructivism, feminist theory, post-structuralism, and post-colonialism. All of these approaches are concerned to illuminate the many ways in which social inquiry is *not* 'value-free' by endeavouring to stand apart from, and reflect upon and transcend, the prevailing order and its dominant perspectives. So, for instance, constructivism, which we discussed in

Chapter 2, sees reality in the social world as something that does not exist independent of us, but that is socially and actively constructed through the process of social interaction. Similarly, post-structuralism maintains that social realities are discursively produced and cannot be understood independent of their dynamic historical construction in social and cultural discourses. It assumes that the way people talk about the world does not reflect some objective truth about the world, but the success of particular ways of thinking and seeing. These ways of thinking and seeing become invisible because they are assumed to be truthful and right. In this way people's thought processes themselves can come to represent and reinforce particular regimes of power and coercion. Hence the purpose of theory is not to provide overarching explanations, or seek causality, but to unpack these discursive processes (see Chapter 14 for further discussion of post-structuralist discourse theory).

Two other critical theoretical approaches seek to reveal, both in social scientific research and writing and in popular discourse, distortions, biases, exclusions, inequalities, and an unacknowledged politics around issues of identity. The first is post-colonialism which, despite its long history in other disciplines, has only recently begun to have an impact in the field of Politics and International Relations. Post-colonial scholars in our field are concerned to illuminate the Western-centric bias of our categories, frameworks, concepts, and historiographic traditions, and how these repress, exclude, marginalize, and objectify the 'other'. They highlight what more mainstream approaches tend to ignore or obscure: that race, nationality, and ethnicity continue to uphold relations of domination throughout the world; and that the Western attempt to judge the world by 'universal' standards such as human rights, free markets, and democracy represents a new form of imperialism.

A second approach, feminist theory, represents a strand of critical theory that has also made a distinctive contribution to our field. Like other critical perspectives, feminist theories open up the field to investigations that go beyond conventional approaches. Its aim is to probe social theories for ways in which gender operates that have gone unnoticed; to trace how gendered boundaries have come to seem neutral and universal—how these boundaries have come to seem to be without gender.

These theoretical approaches are opening up our field—empirically, conceptually, and theoretically—to new avenues of inquiry. In this chapter we address the question of whether social inquiry can be 'value-free' more narrowly. In what follows we first consider how this issue has been addressed in the study of politics. We then consider various ways that 'values' intrude into social-scientific inquiry: how the observer's values and other biases influence research, how the act of observation itself intrudes on the object of study, and the extent to which observation is 'theory-laden', that is, shaped by shared social practices and inseparable from the interpretations to which those practices give rise.

Normative and empirical theory in political research

We can see the influence on political research of the positivist fact-value distinction in the division that has been defined between empirical and normative research and theory. This division is predicated on the assumption that it is possible to separate questions about what is (empirical questions) from questions about what should or ought to be (normative questions).

Positivists insist that only hypotheses about *what is* are constitutive of the domain of value-free **empirical theory**. Thus, according to positivism, it is the ability to keep empirical and normative questions separate that makes it possible for social science to be objective and value-free.

Empirical research and theory is concerned with empirical questions—questions about *what* is—that can be answered with empirical data (data gathered through observations of the world around us). Normative research and theory is concerned with normative questions—questions of *what ought or should be*. It is often associated with questions that address 'moral' issues in politics; with human rights and distributive justice, with what is morally 'just' and what is a just government, and with international order, the distribution of resources, and the way of treating the environment for future generations. When, if ever, is war just, and what is just conduct in war? Positivists insist that questions such as these, questions about what goals we ought to pursue, what direction policy ought to take, or what values ought to be promoted, cannot be addressed by empirical theory and subjected to formal empirical tests.

Research practice tends to reflect this empirical/normative distinction. Researchers tend to address empirical questions by seeking to discover 'facts'; but tend not to undertake 'rigorous, philosophically informed normative study'. On the other hand, researchers address normative questions in ways that do not involve 'rigorous, methodologically informed empirical study' (Gerring and Yesnowitz 2006: 103). Thus, authors of empirical studies will suggest implicitly that their particular object of study 'has a broader significance', that the subject 'affects society', but fail to make explicit how, or to what extent, it actually does so. In addressing normative questions scholars will assume, for instance, that federalism and civil society are good for democracy, or strong parties are good for governance, but will make no attempt to provide empirical proof of the assumption (Gerring and Yesnowitz 2006: 107).

In recent years, critics have challenged this divide between empirical and normative questions in political research. They argue that the notion of 'normative theory' assumes, wrongly, that we can distinguish questions about *what is* from questions about *what ought to be*: they argue that all theory is normative theory. While not all theory is expressly concerned with reflection on normative issues, no theory can 'avoid normative assumptions. Scholars are actually making normative assumptions, though they may not realize it, when they select what data is important, interpret that data, and articulate why such research is significant (Cochran 1999: 1). Robert Cox neatly sums up this overall position: 'all theory is for someone and for some purpose'. All theories reflect values (Cox 1981: 128). If all theory is normative in this sense, then it is impossible to separate facts and values. We begin our exploration of this issue in the next section.

Values, the researcher, and the research process

> [O]ne can do a good job predicting what a study will find by knowing the preferences of the scholars who undertook it.
>
> (Jervis 2002: 188)

The statement quoted above expresses a probably commonly held view: that a researcher's values have at least some influence, not only on the conclusions reached, but on every aspect of research, including the selection of research questions, data gathering, observations,

and conducting of experiments. But even if this is true, does this preclude the possibility of separating facts and values and, therefore, of pursuing objective scientific study of the social world? Some scholars argue that it does; others insist that, while objectivity in the social sciences may be difficult to maintain, it is nonetheless possible.

Max Weber (1864–1920) is among the scholars who have argued for this second position. There are three parts to his somewhat complicated, but highly influential position on this question: (1) There *is* a distinction between facts and values: questions of value are independent of, or separable from, questions of fact. However, he argues that, despite the distinction that can be drawn between values and facts, it nevertheless remains the case that (2) value-neutrality is *not* possible in the social sciences. But, finally he argues that, despite this, (3) we must strive for a value-neutral social science. Let's unpack the implications of this argument by considering each of these points a bit further.

Weber argued that we can distinguish facts from values because knowing the facts of something is not the same thing as knowing its value; and knowing the value of something does not lead us to a factual description. We can therefore also distinguish between *questions* of fact and *questions* involving values. Social scientists should deal only in questions of fact and remain ethically neutral on questions involving values. But he emphasized that values are unavoidable in the practice of social science. Values enter into the selection of problems, the identification of facts, the assessment of evidence, and the content of conclusions. 'There is no absolutely "objective" scientific analysis . . . of "social phenomena" independent of special and "one-sided" viewpoints' (1994: 535).

Let's consider the points raised in this argument. Weber argues that, in the selection of problems to address, researchers select to focus on only a small portion 'of concrete reality'—those 'segments of reality' which are interesting and significant or 'have value-relevance': because they—and only they—are 'related to the *cultural values* with which we approach reality' (1994: 539). It is only those things that we perceive to be meaningful that become objects of investigation for us. Since undertaking research on something presupposes that it is connected in some way to our values, researchers will likely have their own 'point of view' about their chosen subject, and this will influence the conceptual scheme that they construct in order to pursue inquiry. So researchers cannot remain neutral: they inevitably bring their own presuppositions into the research field. There is no 'view from nowhere'.

However, Weber goes on to argue that, though many elements of research are subjective, it is nonetheless possible to conduct research in a value-neutral manner if researchers adhere to the norms and practice of good social science, and if they 'make relentlessly clear', both to their audience and themselves, 'which of their statements are statements of logically deduced or empirically observed facts' and which are evaluative statements (Weber 1948: 2). As long as researchers adhere to scientific standards, then social science can provide a factual and objective assessment of actors and their practices, an objectively true account of how the world functions, without attempting to tell us what goals we ought to pursue, what direction policy ought to take, or what values ought to be promoted.

Weber's views have remained influential and they continue to inform discussions of the role of values in scientific inquiry. Many contemporary scholars follow Weber in arguing that social inquiry can be objective, despite the intrusion of a researcher's own values in the research process; that, while 'important practical obstacles' are frequently encountered

in social science research, that doesn't mean that it is *intrinsically impossible* to secure unbiased conclusions in the social sciences (Nagel 1994: 583).

The philosopher, Ernest Nagel elaborates on the points Weber raises concerning the selection of problems researchers choose to investigate, and the assessment of evidence on which they base their conclusions. With regard to the first point, Nagel, argues that while the selection of problems may be value-orientated, 'It is not clear . . . why the fact that an investigator selects the material he studies in light of problems which interest him and which seem to him to bear on matters he regards as important', presents any greater challenge to the logic of social inquiry than for that of any other branch of inquiry (1994: 572). Nagel cites as an example a researcher who believes that 'a free economic market embodies a cardinal human value' and who wishes to inquire into which activities are important to its operation. Why, Nagel asks, should we assume that his interest in a free market economy will bias his evaluation of the evidence on which he bases his conclusions about what is needed for its operation? The fact that the interests of the researcher determine what they select for investigation 'by itself, represents no obstacle to the successful pursuit of objectively controlled inquiry in any branch of study' (1994: 572).

With regard to the second point Weber raises, Nagel acknowledges that researchers concerned with social phenomena are often committed to social values and that these enter into the assessment of evidence and the content of conclusions. But he argues that 'steps can be taken to identify a value bias when it occurs, and to minimize if not to eliminate completely its perturbing effects' (1994: 573). One way to do this is for social scientists to 'abandon the pretence that they are free from all bias' and 'state their value assumptions as explicitly and fully as they can' (1994: 572). He acknowledges that 'unconscious bias and tacit value orientations are rarely overcome by devout resolutions to eliminate bias'; but there are ways, he argues, to recognize and correct for prejudices. For instance, researchers can 'sift warranted beliefs' and 'retain only those proposed conclusions . . . that survive critical examination by an indefinitely large community of students, whatever their value preferences or doctrinal commitments' (1994: 574). While he recognizes that this mechanism may not work as effectively as in the natural sciences, he insists nonetheless that we are not warranted on these grounds to conclude that reliable knowledge of human affairs is unattainable (1994: 574).

In sum, the position elaborated by Weber and by contemporary scholars like Nagel, is that it is possible to keep our values from influencing our assessment of evidence and to therefore arrive at an objective understanding of the facts of a situation. In Chapter 1 we said that in order to generate valid and reliable knowledge about the questions that we address our process of inquiry, at every step, must be self-aware and critical, and we are required, as researchers to make clear and transparent how our conclusions were reached. We believe that consideration of the issues that we are discussing here will contribute to your ability to conduct research that is critical and self-aware, clear, and transparent.

Consciousness of biases and concerted effort to limit their impact may overcome some of the difficulties involved in separating facts and values. But there are some biases that may be more difficult for researchers to identify and to control.

One bias of this sort is the researcher's *expectations*. What the researcher expects to find can impact the results of an inquiry by influencing, *not* the researcher, but *the behaviour of their subjects*. Researchers call this type of bias the 'Rosenthal Effect', after psychologist Robert Rosenthal. What this involves is explained in Box 3.1, below.

BOX 3.1 The 'Rosenthal Effect'

Rosenthal conducted a series of studies that were designed to determine whether the expectations of researchers can bias their results. In one study, two groups of subjects were shown photographs of people and asked to rate the photographs according to whether the people in them had recently experienced success or failure. The experimenters who administered the test were told what result their group, on average, was expected to produce. Each time this experiment was replicated the results for each group conformed to the expectations that had been communicated to the experimenters. A second study tested the effect of experimenters' expectations on the performance of animals; a third experiment tested the effect of a teacher's expectations on student performance. The results of all the studies showed the same thing: that the expectations of the experimenters influenced the behaviour of their subjects. Further study led to the conclusion that experimenter bias influences *human* subjects through 'subtle and complex auditory and visual cues'; and it influences *animal* subjects through the way they handle and observe them (Martin 1994: 585–7).

So, as Rosenthal learned, a researcher's expectation about the results of an experiment affects the outcome of the experiment. But this type of bias, Michael Martin argues, does not preclude the possibility of value-free inquiry. Experimental techniques that minimize contact between experimenter and subjects, and statistical methods that can correct for bias, mean that the Rosenthal Effect need not foreclose the possibility of objective social science research (Martin 1994: 593).

So, again, despite the evidence that values *do* influence research, it is nevertheless still possible for social science to be objective and value-free. While, there are important practical difficulties in securing value-free and unbiased findings in the social sciences, these are not insurmountable.

We have been discussing how a researcher's biases may intrude into the research process. But the research process can be compromised as a result, not only of a researcher's own biases, but as a result of the research process itself. There are two sorts of biases that illustrate how this might occur. The first is what has been called **going native**. This is where, in the course of pursuing research on a subject, the researcher's interaction with the subject of inquiry—an individual, group, community—creates a bias towards that subject on the part of the researcher. This likelihood can arise in all research, but it has received a good deal of attention in relation to ethnographic research. This is research which explores cultural phenomena from the point of view of the subject of the study (see Chapter 13). Because, in ethnographic research, researchers become immersed in the community they are studying, it may be more difficult for them to remain objective about that community. In some cases, researchers come to identify so much with the subjects who they interact with in the course of studying them that they lose the capacity for objectivity. Again, this source of bias can arise, not only in ethnographic research, but in all types of research.

Let's consider a second example of how the research process may be compromised by biases, not by those that the researcher brings to the process, but by those that arise as a result of the process itself. In this example, the research process is compromised because the behaviour of the subjects we are studying changes as a result of our observing it. While it may be possible to limit or eliminate a researcher's own values or biases, the tendency for the act of observation *itself* to contaminate the subject of study and produce biased and unreliable

conclusions may be a more difficult issue to address. There are a number of related arguments about how this operates.

One argument is that human behaviour changes while we study it, because those we are studying alter their behaviour when they know they are being observed. This phenomenon is known as the 'Heisenberg Effect'. The 'Heisenberg Effect' derives from Werner Heisenberg's Principle of Indeterminacy in particle physics. Heisenberg's principle asserts 'that the very process of trying to observe tiny particles inside the atom alters the situation being studied' (Morgan 1994: 30). Some political scientists have pointed out that the behaviour of government officials and bureaucrats changes when they know they are under observation. This, in fact, is why it is so important in a democracy to have a free press. If politicians know they are being observed by members of the press, they may be deterred (it is hoped) 'from taking actions that would be to their benefit but costly to the public interest' (Morgan 1994: 30).

Here's an example of the 'Heisenberg Effect' on political research. In the course of working as a United States Assistant Secretary of State in charge of education and culture, Charles Frankel discovered that

> when he sent out letters merely to get information about what US officials were doing with regard to particular programmes, he got back replies indicating that the officials to whom he had written had changed what they were doing after receiving his letter.
>
> (Morgan 1994: 37)

So, merely in an effort to inform himself about government programmes, he had 'apparently produced changes in policy', that is, in the phenomena which he was studying (Frankel 1969: 83; quoted in Morgan 1994: 37). In sum, human behaviour changes while, and because, we are trying to study it.

What, then, can we do about this? To deal with this difficulty, researchers use less intrusive methods. For instance, some researchers in our field get around the problem by using 'content analysis' (see Chapter 14). By systematically analysing an official's statements for clues to their perceptions and attitudes, researchers can get material on decision-making without interviewing the decision-makers and so avoid the problem of influencing their data source through the questions they ask. By using texts to collect data, researchers can study past policy positions as they were recorded at the time. Once recorded, texts do not change.

But in addition to altering behaviour as a result of studying it, there is a trickier problem: behaviour may change as a consequence, not only of the process, but of the results, of studying it. The problem here is how we can know if the theories we use to explain outcomes are true if they themselves have the effect of altering the practices that produce those outcomes. An example of this is the effect deterrence theory had on weapons deployments, contingency planning, arms-control proposals, and relations among allies. If theory alters practice, it becomes, in effect, a self-fulfilling prophecy. How then can we claim to be able to separate and distinguish between our observation of the facts, and the theories which are used to test them? Generally speaking, the physical sciences do not face this problem: planets do not move differently because of the process or the results of studying them. But in the social sciences, theory can be a self-fulfilling prophecy. If, using the authority of science, you convince policy-makers that, for instance, war is inevitable given the anarchical nature

of the international system, then it *becomes* inevitable and so it becomes impossible to test the proposition scientifically (Morgan 1994: 31–2).

This problem bears on the question of whether we can separate facts and values because it complicates the notion that there exist mind-independent facts. There are two other arguments concerning how the act or process of observation compromises our ability to separate facts (observations) and values (theories). Perhaps the best way to make clear their substance and implications is by briefly revisiting Karl Popper's (1959) critique of classical positivism, which we discussed in Chapter 2. Recall that Popper's critique involved two arguments. The first is that, rather than endeavouring to discover laws through induction, scientists should use a process of deduction which starts with a general or logical proposition and then 'tests' this theory with observations derived from it. His second argument is that falsification, not verification, should be the aim of science; that the aim of our tests must be to 'falsify', rather than verify, them. We reject those which are falsified but continue to test those that are not until they become so thoroughly tested that we can consider them to be 'confirmed', though it remains possible that someday someone may falsify or significantly modify them.

Popper's first argument assumes that the observations or facts we use to test theories can be established independently of the theory that they are meant to test. But Thomas Kuhn (1962) and others have argued that observation is 'theory-laden': that our observation of 'facts' cannot be separated from the theoretical notions which we bring, and which give intelligibility, to what we observe. Observation is the interpretation of a phenomenon in the light of some *theory* and other background knowledge. In fact, we might say that a phenomenon is an already interpreted regularity (or event). An interesting way of illustrating this, is through the famous duck/rabbit picture shown in Figure 3.1. As Kuhn explains (1962: 114), when we look at this picture we do not merely observe curved lines on a page: we see either a rabbit or a duck.

Can we perceive what these lines produce irrespective of an act of interpretation? While the same data lend themselves to different interpretations, the perceptual experience and theoretical interpretation are simultaneous: the 'facts' do not present themselves to us independently of our theoretical beliefs. What we observe this object to be is dependent on how we describe it. That is, if we are told that 'this is a picture of a duck', we will likely see a duck; if we are told that it is a rabbit, we will probably see a rabbit. The reality is not independent of a description of that reality.

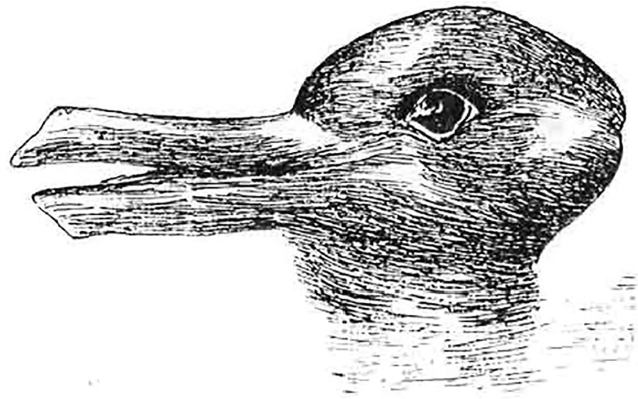


Figure 3.1 Duck or rabbit?

Source: J. Jastrow (1899).

Popper's second argument, that falsification rather than verification should be the aim of science, highlights another problem with the fact (observation)—value (theory) distinction. Popper assumes that the scientific enterprise involves the rigorous, systematic application of reason, so that when theories are falsified by experience they are thrown out, irrespective of how many important people have invested themselves and staked their careers in promoting them. But some people question whether this assumption is warranted. Is this actually the way scientists go about developing and testing theories? Is it actually the case that researchers seek to disprove or falsify their own theories? *Do* they discard their theories when they are confronted with disconfirming evidence?

The philosopher of science, Thomas Kuhn, argued that Popper (1959) presents an idealization of true science; that, rather than unrelenting criticism and testing, science has tended towards paradigmatic conformity and conservatism. In *The Structure of Scientific Revolutions* (1962), Kuhn argues that science is a social institution. It consists of a community within which there is a common view—what Kuhn calls a **paradigm**: a conceptual scheme, about which some community of scholars agree, and which defines the objects of investigation and the ways in which they are to be investigated. The implications of this argument for the fact-value distinction and, thus, for the objectivity of scientific knowledge, has made Kuhn's book a focus of controversy for over half a century. We pursue this in more detail, below.

Values and social practice

So far, the discussion has focused on how a researcher's values intrude on the research process. But to adequately address the question of how values influence social research we need also to consider the influence of values associated with a more general set of shared social practices.

Those who argue that is it possible to conduct value-neutral social scientific research tend to assume that social inquiry is unaffected by the context and environment in which it is conducted. But this ignores the fact that researchers live within a given society, and work within a system of academic incentives and disincentives that reward some activities and discourage others. It would seem unreasonable to suppose that this system, including the gatekeeping functions of the academy, the creation and enforcement of disciplinary norms, 'rewarding good behaviour and punishing—or at least ignoring—bad behaviour' (Gerring 2001: 5), public and private funding, publication outlets, universities, government, and the mass public, *does not* affect researchers and their work. The question is whether such influences invalidate the possibility of a neutral social science. In the next section, we discuss two influential perspectives on this question. We start by considering Thomas Kuhn's influential arguments concerning paradigms and how they change through scientific 'revolutions'. After reflecting on their implications for the possibility of value-free social inquiry, we then explore the implications of Imre Lakatos' rival notion of 'scientific research programmes'.

Thomas Kuhn and scientific revolutions

The publication of Thomas Kuhn's book, *The Structure of Scientific Revolutions* unleashed a firestorm of controversy which continues to reverberate today. Kuhn (1922–96) is widely considered to have been 'one of the most influential philosophers of science of the twentieth

century, perhaps the most influential', and his book, *The Structure of Scientific Revolutions*, is 'one of the most cited academic books of all time' (Bird 2009). By highlighting the value-laden social practices within which scientific inquiry takes place, the book presents a compelling challenge to the notion that it is possible to separate facts and values in scientific research.

In *The Structure of Scientific Revolutions*, Kuhn surveys the history of science and, based on this history, concludes that the growth of scientific knowledge is not a *logical* process but a *social* one. Science, he claims, does not progress by stages based upon neutral observations and through cumulative, gradual knowledge acquisition, but through an essentially *social* process. He argues that science is a social institution in which scientists are socialized, not to pursue unrelenting criticism and testing, but to accept the reigning values, beliefs, concepts, and rules of their profession. Consequently, there is no 'theory-independent' view of the world: instead, the world appears to us in the context of theories we already hold: it is 'theory-laden'.

To capture this view of science, Kuhn introduces the concept of a 'paradigm'. For Kuhn, a 'paradigm' is a conceptual scheme, about which some community of scholars agree, and which defines the objects of investigation and the ways in which they are to be investigated. Science, Kuhn argues, tends to be dominated by 'paradigms'. Once a paradigm is established, it directs scientific investigation through the operation of what Kuhn calls 'normal science'. The aim of 'normal science' is not 'to call forth new sorts of phenomena', but 'to force nature into the preformed and relatively inflexible box that the paradigm supplies'; indeed 'those that will not fit the box are often not seen at all' (1962: 24).

Though Kuhn's book was concerned with the natural sciences, it caused a great crisis of objectivity in both the natural *and* the social sciences. The crisis arose because Kuhn's arguments threw open to question the status of science as a rational and progressive development towards 'truth', challenging the distinction between natural and social science, and between social science and ideology.

Positivists maintain that if the study of politics is to be scientific, it must adopt the same scientific principles that the natural sciences use to study natural phenomena. But Kuhn challenges the notion that the natural sciences are objective or value-free; that scientists apply the neutral instrument of 'scientific method' to independently existing reality; that the growth of scientific knowledge is piecemeal and cumulative; and that theories are discarded as 'falsified' if they fail (Ball 1976: 158). Based on his reading of the history of science, he argues that the validity of truth-claims is grounded in the consensus of some scholarly community; that what science means by 'truth' is the consensus of those it defines as competent and full members of its community. Academic journals, hiring practices, and curricula are orientated towards certain explanatory frameworks and conceptual schemas, and research using other frameworks and schemas is undervalued and marginalized within that community.

These arguments challenge the distinction in science between fact and value and, consequently, our faith in the authority of 'science'. Science is essentially social. It consists of a community within which there is a common view. This 'common view' is what Kuhn calls a 'paradigm'.

Paradigms and paradigm change

The Structure of Scientific Revolutions argues that the history of science can be described in terms of 'paradigm shifts'—transitions from one paradigm to another. These shifts occur when an anomaly or anomalies arise which violate the assumptions of the reigning

paradigm. But the new paradigm which emerges is not the result of a rational and progressive development towards 'truth', but of a scientific 'revolution' which, in its essential characteristics, bears an uncomfortable resemblance to a social or political revolution.

The process through which science develops can be described, then, as one involving the consolidation of paradigms, and processes which shift the commitment of a community of scholars from one paradigm to another. Kuhn develops a model which elaborates this. The model starts by describing the features of what Kuhn calls *pre-paradigmatic scientific investigation*. This is a pre-scientific phase in the development of a scientific field in which, due to the absence of a paradigm, investigation is random and diverse. The absence of a paradigm or some candidate for one, means that 'all of the facts that could possibly pertain to the development of a given science are likely to seem equally relevant'. As a result, 'fact-gathering' is an activity that is far more 'random' than the one which occurs once the establishment of a paradigm inaugurates what Kuhn calls 'normal science' (1962: 15).

While we are in this pre-paradigmatic period of investigation, there will be a variety of schools of thought that compete for our attention. But, in what the model describes as the next phase in the development of science, one of these pre-paradigmatic schools triumphs and this leads to the *establishment of a paradigm*. It is then that scientific investigation ceases to be random and becomes highly directed and focused. This is because paradigms generate 'particular coherent traditions of scientific research', reflected in textbooks, departments, degrees, appointments, promotions, journals, conferences, and honours and awards (1962: 10). Consequently, with the establishment of a paradigm, we enter into a period of what Kuhn calls '*normal science*': science that, operating within the limits of a paradigm, concerns itself with testing and working on 'puzzles' via bits of research that are implied by the paradigm. Scientific research in a period of normal science, consists of developing the reigning paradigm 'by extending the knowledge of those facts that the paradigm displays as particularly revealing, by increasing the extent of the match between those facts and the paradigm's predictions, and by further articulation of the paradigm itself' (1962: 24). This involves both empirical and theoretical work. The empirical work consists of fact gathering, improving the accuracy and scope of factual determinations, matching facts with theory, coming up with quantitative measures, and applying the paradigm to new areas and different settings. The theoretical work involves using the theories generated by the paradigm to predict, and developing new or more precise applications of them. But whether empirical or theoretical, the work of normal science does not raise questions about the paradigm itself: it treats the paradigm as a given and does not pursue critical investigation of its core assumptions. The assumption on which normal science is based, and which scientists aggressively defend, is 'that the scientific community knows what the world is like'. Consequently, the work of normal science is devoted to 'a strenuous and devoted attempt to force nature into the conceptual boxes supplied by' the paradigm (1962: 5).

However, it *is* possible for research within the framework of the paradigm, or for real-world events and processes, to bring to light anomalies for the paradigm. An *anomaly* is something which deviates from the general rule; something which is inconsistent with its basic assumptions. But anomalies are not grounds for rejecting a paradigm: they generally do not even cause a critical review of it. However, if the anomalies grow, or if an anomaly is very fundamental—if it threatens to undermine the very foundation of the paradigm—this may lead to a *crisis*. An anomaly may also evoke a crisis if it has some immediate practical import. In this situation, the field starts to revolve around the problem and to search for its

solution, and the field may consequently begin to look different. However, scientists resist change. They will set the problem aside, or devise modifications in order to shoehorn the anomalous fact into the existing paradigm. And even when the crisis begins to blur the paradigm—to loosen the rules of normal research in the field—theorists will not declare the crisis-ridden paradigm invalid unless there is some alternative paradigm available to take its place.

Once a candidate for a new paradigm emerges, there is a battle over its acceptance; and, finally, there is a *transition* from the paradigm in crisis to the new one from which a new tradition of normal science emerges. This constitutes a revolution in the field. As in a political revolution, in a 'scientific revolution':

1. A malfunction leads to crisis;
2. There is a loosening of rules as basic institutions are rejected;
3. Competing camps form: there are those who seek to defend the old order, and those who seek to institute something new;
4. The competing camps seek support by resorting to techniques of mass persuasion;
5. Scientists transfer their loyalties from the old to the new paradigm, and this inaugurates a new era of normalcy or 'normal science'.

Figure 3.2 depicts these steps and how they are related.

Consider the implications of this model for the way we understand scientific claims! Kuhn's notion of 'normal science' suggests that the rationality of science is limited. You may recall that in Chapter 2 we said that, for positivists, the distinction between facts and values is what differentiates science from ideology; that, unlike science, religions and ideologies are non-falsifiable because, rather than proving them false, potentially disconfirming or

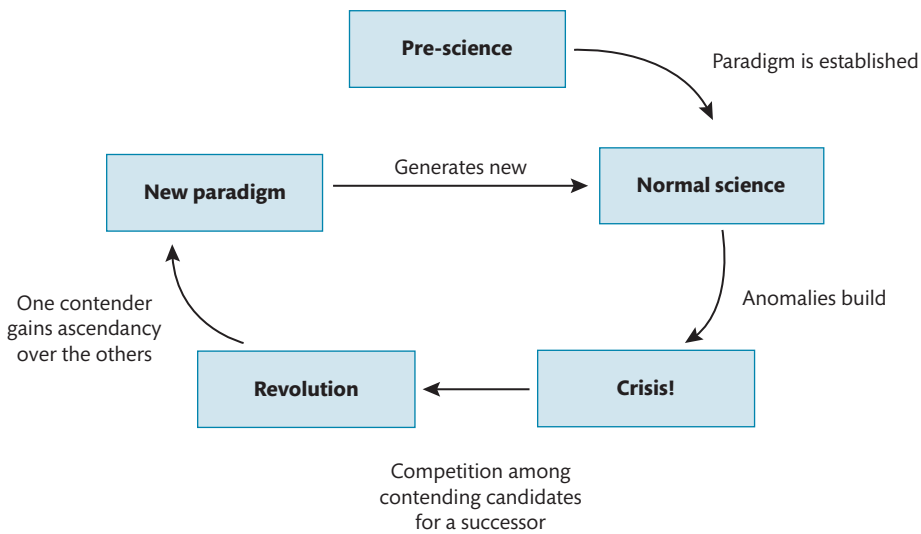


Figure 3.2 Thomas Kuhn, *The Structure of Scientific Revolutions*

anomalous facts are incorporated within them. Kuhn suggests that, just as adherents of an ideology or religion will discard an anomaly or disconfirming fact or incorporate it within the ideology, scientists will 'often suppress fundamental novelties because they are necessarily subversive' of their 'basic commitments' (1962: 5). Similarly, the shift from one paradigm in favour of another is not the result of a rational decision but something rather more like a conversion experience. Paradigms shift when scientists change their minds about fundamental matters.

This represents a clear challenge to the positivist ideal of scientific progress. Positivists see the choice between two competing theories as resolved by a 'crucial experiment' whose results support one theory and refute the other. But, for Kuhn, the shift from one paradigm to another neither results from nor produces cumulative knowledge because different paradigms produce *incommensurable* knowledge. Paradigms are incommensurable because they disagree about what the data actually are; so paradigm choice can never be settled by logic and experiment alone. It involves, at least in part, an act of faith.

In sum, a paradigm generates, or consists of, a set of theories, that not only shares common assumptions, but whose assumptions are incommensurable with respect to those of some other paradigm. This is because paradigms do not contain theory-neutral facts that can serve as a basis for comparison. All observation is theory-laden: 'When Aristotle and Galileio looked at swinging stones, the first saw constrained fall, the second a pendulum' (Kuhn 1962: 121). They relied on such different ontological assumptions that they were incommensurable; so no neutral measure could be found to assess one against the other. The arbiter, Kuhn argues, is peer consensus. And the new 'paradigm', or way of looking at the world, may not contain any of the ideas of the old one.

Kuhnian paradigms and political studies: the case of development theory

As previously noted, Kuhn's model of scientific progress was developed to describe the natural, not the social, sciences. Kuhn considered the *social* sciences to be 'pre-paradigmatic' and unlikely to ever become, as in 'mature sciences', capable of the paradigm-driven investigation that characterized the natural sciences (Kuhn 1962: 164–5). However, many social scientists saw Kuhn's elaboration of the role of paradigms in scientific development 'as an extremely suggestive description of their own disciplines' (Gutting 1984: 1). Consequently, the idea that paradigms direct and shape research quickly became 'widely accepted in the social sciences' (Bernstein 1981: 425). The 'profound influence' it is has had on Politics and International Relations is clear to see (Ball 1976: 151). In our field, the term 'paradigm' has been applied to specific methodological and theoretical schools of thought such as behaviouralism and rational choice theory, and scholars have devoted considerable energy to establishing, and defending, paradigms in International Relations and Comparative Politics (Schmidt 2002: 9).

The term 'paradigm' is defined in a variety of ways, even in Kuhn's work, so it is worth clarifying how scholars in our field use the term. In the study of politics, the term is applied to world views or schools of thought that provide a theoretical architecture for pursuing the study of the political world. Each paradigm consists of a set of assumptions that provide a basis for the development of theories to explain particular events or categories of events and that enable us to decide, from among the universe of political phenomena, which are

important to study. Different behaviours and prescriptions for action flow from different sets of assumptions. If you assume people are capable of altruism, mutual aid, and collaboration, you might be in favour of pursuing cooperation and collective action to resolve problems and advance global welfare. If you assume that humans are driven, first and foremost, by an instinct for power, your choices might be more fatalistic and survivalist. If you assume that people are basically a product of the material conditions and social relations which dominate their life, then, if you wish to change the world, you may seek to change the material environment which shapes human thought and action. This general usage of the term 'paradigm' in the study of politics is summarized in Box 3.2.

Let's consider an example of how the notion of 'paradigm' has been applied to a specific area of inquiry within politics: the study of development.

Paradigms and the study of development

Since the Second World War, the field of development studies has been dominated by two theoretical perspectives—'modernization theory' and 'dependency theory'—which are frequently referred to as 'paradigms' (Bodenheimer 1971; Foster-Carter 1976; Valenzuela and Valenzuela 1979; Pavlich 1988; Geddes 2003). Barbara Geddes has argued that, while neither perspective has operated as a 'Kuhnian hegemonic paradigm', they have 'most of the other features Kuhn attributes to paradigms' (2003: 6–7). Each is comprised of 'collections of theories, hypotheses, applications, and favoured methodologies' (2003: 21). In each, a set of theories structures research, including identifying what questions need to be addressed and what constitutes evidence. Modernization theory is based on Liberal and neo-Liberal assumptions, while dependency theory is based on Marxist and neo-Marxist assumptions. Each consequently describes and explains 'third world' underdevelopment in strikingly different ways.

BOX 3.2 Paradigms in Political Research

Paradigms

- arise within a specific social context
- consist of a set of assumptions about the nature of an area of study
- help to explain events and processes, and provide a guide to action or policy
- lose their usefulness when events or circumstances arise which violate the assumptions of the paradigm, and are supplanted by other paradigms, containing different assumptions

During the immediate post-Second World War decades, a paradigm for the study of development emerged in the United States which reflected the social context of that place and time. US foreign policy following the Second World War was concerned to ensure that the 'developing world' would not be drawn into the Soviet communist bloc; and it was this

concern that animated the US government to enlist social scientists to study and devise ways of promoting capitalist economic development and political stability (Gendzier 1985; Diamond 1992; Simpson 1999). By the early 1950s, the study of 'development' was established as a key scholastic project and, through generous funding and institutional inducements, it continued to attract a steady and ever-expanding flow of research and writing from across the social sciences. Within a couple of decades, the study of development had markedly converged around a common set of analytic conventions and general themes. These conventions and themes eventually became the basis for what is referred to as 'modernization theory'.

Modernization theory assumed that underdevelopment was a stage in development through which all nations must pass, and that progress would eventually come about through the spread of modern capitalism to backward, traditional areas. But, consistent with strands of neo-Liberal thought current in post-war America, it maintained that development was not something that could be achieved through more-or-less natural processes of economic growth, but only through the intervention of more advanced countries. Problems of development, it maintained, result from weaknesses in the various factors of production—land, labour, and capital—within the developing countries; and the way these weaknesses could be overcome is through trade, foreign investment, and aid from advanced capitalist countries. Ultimately, and through contact with the developed market economies, the diffusion of capital and technology would produce economic development in the backward countries.

However, by the end of the 1960s, a key anomaly had emerged for the modernization paradigm: 'modernization' did not appear to be occurring in 'developing countries'. The paradigm had generated the expectation that the development of industrial capitalism in these countries would enable them to achieve the conditions of life that exist in the advanced industrial countries of 'the West'. However, it soon became clear that converging levels of industrialization had not produced a concomitant convergence of incomes; and that the divide in income and wealth between the 'developed' and 'developing' worlds was, in fact, widening. For whole societies and large populations within them, industrial 'development' appeared to be producing conditions of life that were worse and not better.

In the late 1960s, a theoretical perspective emerged that offered an explanation for this anomaly. The overall conception it advanced was of capitalist development as a worldwide process that, as the result of the appropriation of surplus by advanced countries from those in the 'developing' world, delivers unequal benefits and produces structurally different developmental outcomes. This perspective formed the basis of what came to be known as 'dependency theory'.

Dependency theory not only challenged many of the core assumptions of modernization theory, it turned modernization theory on its head. So, for instance, while modernization theorists focused on the benefits of free trade, foreign investment, and foreign aid, dependency theorists argued that international market relations occur in a framework of uneven relations between developed and underdeveloped countries and work to reinforce and reproduce these relations. Dependency theorists argued that the international economy is under the monopolistic control of the developed economies; and that free,

unregulated international trade and capital movements therefore tend to favour the already well-endowed and work to accentuate, not diminish, international inequalities. They argued that a more equitable distribution of benefits could not occur within the international capitalist system, and that 'developing' countries would therefore be unable to move beyond limited industrialization.

But, as modernization theorists and other critics soon pointed out, there had emerged a key anomaly for dependency theory: the demonstrated capacity of some 'third world' countries to industrialize and achieve rapid economic growth. Dependency theorists responded by insisting that, though the manufacturing sector in many 'developing' countries had grown dramatically, these sectors were controlled by multinational corporations with headquarters in the advanced countries; thus, whatever benefits they might bring in the form of managerial and technological know-how, it remained the case that they take more than they give and make it impossible for these countries to achieve the conditions of life that exist in the advanced countries. While critics acknowledge that many problems of development persist in countries that have achieved substantial industrial growth, they argue that the explanation that dependency theorists offer for them is unconvincing, that it fails to sufficiently account for different developmental outcomes, vastly exaggerates the power of the international system to shape outcomes in developing countries, and systematically ignores the role of internal factors in producing and reproducing these problems.

The debate remained at a stalemate. Adherents of both perspectives failed to undertake the sort of vigorous criticism, empirical testing, or theoretical investigation that might open the way to radical revision, synthesis, or consensus. Instead, modernization theorists and dependency theorists continued to assert their commitment to the core assumptions of their respective paradigms: one set of theorists continually asserted the necessity for greater market freedom; another set continued to focus on the enduring legacy of colonialism and the continued relevance of imperialism. With development theory at an impasse, many theorists simply undertook to reduce expectations of development, so that, by the second half of the 1990s,

> the normal usage of the term 'development' had mutated to mean a modest increase in 'industrialisation' . . . an improved capacity to produce textiles, sports shoes, furniture, or even a solid GDP based on the export of resources, agricultural products, or low-tech industries.
>
> (Mason 1997: 409)

To many observers, the field of development studies appears to have been driven, less by the findings of rigorous scientific investigation, than by incommensurable and irreconcilable ideological commitments. Barbara Geddes has argued that inquiry into development has been characterized by the succession of one untested theory after another, bringing us no closer to truth but, instead, embroiling us 'in fruitless, often ideologically weighted debates' (2003: 223). The problem with the two perspectives, she argues, is not so much their incapacity to resolve anomalies, as the paradigmatic conformity that led researchers to ignore disconfirming evidence. In each paradigm, 'evidence was available that should have called them into question at the time of their creation'. Yet, it took 'decades to notice the plethora of inconvenient facts' (2003: 7–8), because researchers failed 'to make use of all

available information' in the formulation of theories, and consumers were willing 'to accept theories without strong supporting evidence' (2003: 17). Analysts used evidence selectively 'to develop, support, or test theories': the modernization paradigm 'used evidence from only a few North Atlantic countries'; while earlier versions of dependency theory 'ignored readily available evidence of rapid growth' in developing countries' (2003: 17–18). Dependency theory was revised in response to evidence of industrial growth in developing countries. However, 'cross-national studies that challenged the implications of the revised dependency paradigm were ignored' (2003: 15). She notes that scholars find persuasive and are drawn to a theory because of its intuitive, emotional, and ideological appeal. They prefer one theory over another less for its explanatory abilities than because of its consistency with their own interests, beliefs, values, and personal morality; and when a theory fits with their personal experience or ideological commitments, they are less inclined to dig deep for disconfirming evidence. What researchers see is essentially related to their theories and beliefs. Available information that is inconsistent with a researcher's preconceptions is disregarded. In this way, paradigms become, in effect, a means of excluding other interpretations and their policy implications.

These reflections are consistent with Kuhn's notion of science as 'social', as is Geddes' account of the shift from one developmental model to another. Geddes observes that 'Each paradigm lost its power to structure research when evidence of gross inconsistency between expectations arising from theories within the paradigm and real-world events became impossible to ignore' (2003: 7–8). When this occurred, modernization theory was overthrown by dependency theory in the same way that political regimes are toppled: by 'well-organized, coherent, mobilized opposition'. The subsequent demise of dependency theory was due to its own 'internal contradictions and inability to deal with the inconvenient facts thrown up by the world'; and, like the collapse of a regime, the fall of the dependency paradigm was followed by a period of 'chaos and contention' (2003: 7).

While Kuhn's arguments continue to provoke reflection on the nature of the social-scientific enterprise, a second, contrasting conception of the nature of science and how it advances has also proved highly influential. But, before turning to a consideration of this second perspective, it is worth noting a position that takes seriously Kuhn's notion of 'normal science' and calls for a 'post-normal science'.

The notion of 'post-normal science' was advanced by Silvio Funtowicz and Jerome Ravetz (1992, 1993) to describe an approach that is complementary to, but different from, conventional or 'normal' science. They argue that, while 'normal science' is appropriate in situations with low levels of uncertainty and risk, it is not suitable in situations when either decision stakes or system uncertainties are high. With low levels of uncertainty and risk, the standard techniques and procedures of scientific inquiry, and the peer review process for ensuring the quality and validity of results, remain appropriate. But Funtowicz and Ravetz argue that 'traditional problem-solving strategies of science . . . need to be enriched to solve the problems that our science-based industrial civilization has created'. They call for a 'democratization' of science that entails, not only incorporating multiple viewpoints into the process of inquiry, but extending the peer community. Their notion of post-normal science is summed up in this excerpt:

> The problem situations that involve post-normal science are ones where, typically, facts are uncertain, values in dispute, stakes high, and decisions urgent. Because applied science and professional consultancy are inadequate, something extra must be added onto their practice which bridges the gap between scientific experts and a concerned public. This is post-normal science, comprising a dialogue among all the stakeholders in a problem, regardless of their formal qualifications or affiliations. For the quality assessment of the scientific materials in such circumstances cannot be left to the experts themselves; in the face of such uncertainties, they too are amateurs. Hence there must be an *extended peer community*, and they will use *extended facts*, which include even anecdotal evidence and statistics gathered by a community. Thus the extension of the traditional elements of scientific practice, facts, and participants creates the element of a new sort of practice. This is the essential novelty in post-normal science. In this way we envisage a democratization of science, not in the sense of turning over the research labs to untrained persons, but rather bringing this relevant part of science into the public debate along with all the other issues affecting our society.
>
> (Funtowicz and Ravetz 1992: 254–5; emphasis in original)

While this perspective shows how Kuhn's arguments continue to inform thinking about the nature of the social-scientific enterprise, it leaves unanswered questions about the research methods, and processes of validation, that are appropriate for post-normal science. Let's turn, then, to a contrasting conception of the nature of science.

Imre Lakatos and scientific research programmes

In his essay, 'Falsification and the Methodology of Scientific Research Programs' (1970), the philosopher of science, Imre Lakatos (1922–74), emphatically rejects Kuhn's model of scientific progress. Kuhn's model, he argues, is relativistic and reduces science to 'mob-psychology' (1970: 178). Lakatos' aim, therefore, is to elaborate a model of scientific change that highlights how scientific change involves, at every step, a process that is both critical and rational. He replaces Kuhn's 'paradigm' with the notion of a '*scientific research programme*', and describes how the incremental, cumulative, and progressive articulation of scientific research programmes lead to the growth of scientific knowledge. He rejects the notion that there is anything like 'normal science' where everyone agrees on the basic assumptions defining a subject of investigation and the way in which it is to be investigated. He argues that there are always alternative, competing theories in existence. What Kuhn calls 'normal science', he argues, is nothing but a research programme that has achieved monopoly; but research programmes rarely achieve complete monopoly and, when they do, they do so only for relatively short periods (1970: 155).

A 'scientific research programme' consists of two main elements: the 'hard core' and the 'protective belt'. The relationship between these two elements is defined by two methodological rules, which he calls the 'positive heuristic' and the 'negative heuristic'. Box 3.3 provides a definition of these terms, along with a number of other terms relating to them.

The *hard core* of a research programme consists of very general hypotheses. It is a 'hard' core because of its *protective belt* of auxiliary assumptions. The *negative heuristic* of a research programme is a convention or methodological rule that stipulates that scientists should not question the hard core of the programme, that they should not modify or abandon it, and should avoid paths of research that bear directly on it. The

BOX 3.3 Imre Lakatos and the Methodology of Scientific Research Programmes: Basic Terms

Scientific research programme	The development of a theory, which remains constant (the 'hard core'), in combination with its protective belt of auxiliary hypotheses, which changes over time.
Hard core	Basic axioms and hypotheses that remain constant and are accepted (for a time) without question.
Protective belt	Specific falsifiable theories or 'auxiliary hypotheses' based on the hard core; these, but not the hard core, can be falsified.
Auxiliary hypotheses	Additional specific hypotheses generated by a theory.
Heuristic	A rule or recommendation for developing an idea or theory.
Negative heuristic	The stipulation that the hard core of the programme not be abandoned or modified.
Positive heuristic	Instructions as to how to adjust the theories and auxiliary hypotheses in the protective belt in the face of an anomaly.
Anomaly	A departure from what is expected or predicted in a specific set of circumstances.
Progressive research programme	When adjustments to the protective belt predict hitherto unexpected facts, the research programme is *theoretically progressive*; it is *empirically progressive* if some of these predictions are corroborated.
Degenerating research programme	When adjustments to the protective belt are made that fail to explain and corroborate new facts.

positive heuristic tells scientists to pursue research related to the hypotheses that make up the protective belt; and provides general guidelines for changing or augmenting them in the face of an anomaly. The protective belt can be modified or changed without abandoning the programme itself, so when we get a falsifying observation, we should attribute it to problems with auxiliary hypotheses that constitute the programme's protective belt. We are free, in fact encouraged, to question, modify, and change these hypotheses in order to protect the hard core.

A research programme changes, then, by modifying auxiliary assumptions and hypotheses in the programme's protective belt. These are tested, adjusted and readjusted, and even completely replaced. As Figure 3.3 shows, this process may produce a problem-shift. Lakatos was deeply hostile to Kuhn's argument that paradigm shifts depend ultimately on the consensus of the scientific community. He saw this as tantamount to a claim that scientific change is non-rational, essentially a process of arriving at 'truth by consensus' (Lakatos 1970: 92). Consequently, his concern was to show that there are objective reasons for why scientists favour one theory, or research programme, over another; that the process by which we choose from among competing research programmes is a rational process (and not a political battle, as is suggested in Kuhn's account).

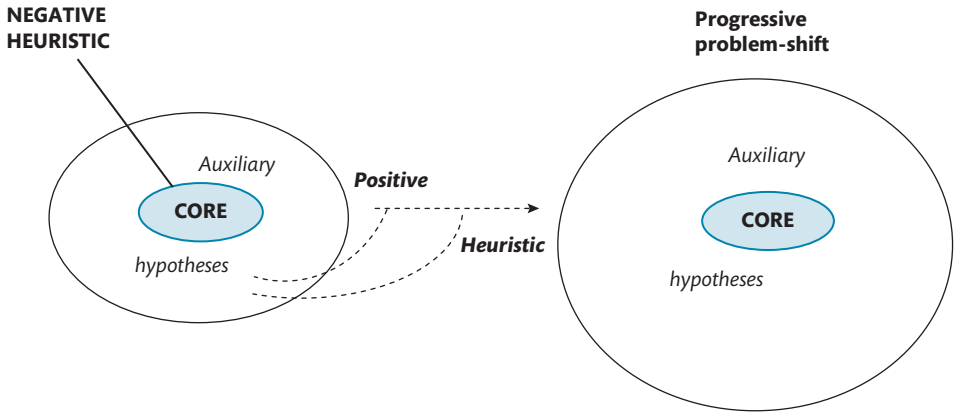


Figure 3.3 Imre Lakatos' methodology of scientific research programmes

A progressive problem-shift leads to the discovery of novel facts. If a research programme increases in content through progressive problem-shifts, it is progressive; if it fails to do this, it is degenerating. Here is how Lakatos describes these processes:

> Let us take a series of theories, T1, T2, T3, . . . where each subsequent theory results from adding auxiliary clauses to . . . the previous theory in order to accommodate some anomaly, each theory having at least as much content as the unrefuted content of its predecessor. Let us say that such a series of theories is *theoretically progressive* (*or 'constitutes a theoretically progressive problem-shift'*) if each new theory has some excess empirical content over its predecessor . . . Let us say that a theoretically progressive series of theories is also *empirically progressive* (*or 'constitutes an empirically progressive problemshift'*) if some of this excess empirical content is also corroborated . . . Finally, let us call a problem shift *progressive* if it is both theoretically and empirically progressive, and *degenerating* if it is not.
>
> (Lakatos 1970: 118; emphasis in original)

In general, we can assess the relative merits of competing research programmes by comparing the degrees to which they are progressive or degenerating—the degree to which the problem-shifts through which each is developing are progressive or degenerative. Again, and as Figure 3.4 shows, we do this by determining whether the problem-shift leads to the acceptance of a theory which has produced a higher ratio of novel facts to anomalies.

Theory #1		**Theory #2**
Ratio of novel facts to anomalies	**<**	**Ratio of novel facts to anomalies**

Figure 3.4 Imre Lakatos' methodology of scientific research programmes: comparing theories

Research programme A is preferable to competing research programme B if either A is more *progressive* than B or is *degenerating* less than B. Although not everyone will come to reject a research programme at the same time, if the changes made to the protective belt to save the hard core from falsifying observations are continually ad hoc and do little to increase the scope of the research programme, the research programme as a whole will eventually be rejected.

In sum, science is both progressive and rational. The methodology of research programmes leads to scientific theories that get closer and closer to the truth. The decision to abandon one research programme in favour of another is not non-rational (as Kuhn appears to assert); it is at every step a critical, considered decision in the light of available alternative theories against which the progressiveness (or degeneration) of successive problem-shifts is gauged.

Like Kuhn, Lakatos was writing about the natural sciences. But his notion of scientific progress through the articulation of 'scientific research programmes' was soon extended to the field of politics. So let's consider how a Lakatosian perspective describes the development of post-Second World War development theory.

In his survey of the history of post-Second World War development theory, Richard Higgott argues that the two dominant approaches to development—modernization theory and what he refers to as 'Marxist' theory ('broadly defined')—are best understood, not as paradigms, but as Lakatosian 'research programmes'. While the 'study of underdevelopment has been marked by a polarization into [these] two broad based schools of thought' (1983: 14), neither of them is 'tightly-paradigmatic' in the sense popularized by Kuhn (Higgott 1983: 43). Neither achieved 'an ascendancy which would allow for a process of normal and revolutionary science' (1983: 6). Moreover, there was no 'paradigm shift' from one to the other.

Recall that, in Barbara Geddes' account, *modernization theory* had been 'toppled' and replaced by dependency theory (which most scholars associate with Marxist theory, 'broadly defined'); that, subsequently, *dependency theory* had collapsed due to its own 'internal contradictions and inability to deal with the inconvenient facts'; and that this collapse was followed by a period of 'chaos and contention' (Geddes 2003: 7). But Higgott insists that the crisis within modernization theory led, not to a paradigm shift, but to internal restructuring. Consequently, with the rise of dependency theory, there was no 'wholesale migration of scholars' from one to the other (1983: 9), no 'transference of allegiance, or conversion', such as occurs in a Kuhnian paradigm shift (1983: 42). Nor, in Higgott's account, did dependency theory collapse and inaugurate a period of 'chaos and contention'. According to Higgott, dependency theory arose in the 1960s as a 'crude radical alternative' to modernization theory (1983: 5), but it exhibited 'highly pluralistic tendencies' (1983: 46), and ultimately served as a springboard to a richer, more sophisticated Marxist analysis (1983: 52).

Consequently, neither perspective 'was killed off (or falsified)'. Rather than one succeeding the other, the two traditions co-existed and, not only endured, but prospered. At times they seemed 'at risk', especially in the 1960s and early 1970s when testing 'appeared to be degenerative or "content decreasing"'. But both proved to be 'extremely durable' because 'the essence of these research programmes' was 'basically sound'; and the modifications introduced within each perspective were progressive adjustments to their theoretical hard cores (1983: 8).

The role of values in research: Kuhn and Lakatos compared

It was Lakatos' intention to elaborate a contrasting model to the one Kuhn had developed. However, in their respective accounts of how science actually operates, the two models appear more similar than dissimilar. Both assume that scientists work within a framework of theoretical assumptions; and that scientists maintain a framework's central set of propositions by convention. To this extent, it seems right to argue, as Kuhn does, that the Kuhnian notion of 'paradigm' and the Lakatosian conception of 'research programme' describe the same phenomenon. Kuhn argues that:

> in discussing research conducted within a tradition . . . I have repeatedly insisted that it depends, in part, on the acceptance of elements which are not themselves subject to attack from within the tradition and which can be changed only by a transition to another tradition, another paradigm. Lakatos, I think, is making the same point when he speaks of the 'hard core of research programs,' the part which must be accepted in order to do research at all and which can be attacked only after embracing another research program.
>
> (Kuhn 1970: 137)

But, while both agree that scientists maintain a tradition's core assumptions, they offer different explanations for why scientists do this. Kuhn argues that the defence of a research tradition's core assumptions is often due to a tendency on the part of scientific communities to encourage paradigmatic conformity; Lakatos disagrees: he insists that this defence is the outcome of rational methodological decisions. Both invoke the history of science to support their claims. Based on the historical evidence they present, it could be argued that Kuhn's evidence suggests that the commitment of a scientific community to a research tradition does not always have an entirely objective basis. On the other hand, Lakatos does not really attempt to show that, historically, it is actually the case that scientists direct criticism away from the hard core of a research programme because they make a methodological decision to do so. It would seem, then, that while Kuhn's argument is concerned with what, historically, scientific practice has been, Lakatos is making a normative argument about what scientific progress *should* entail.

Kuhn points to what he sees as a second way in which the two models appear to be similar: the description of what Lakatos calls the 'degenerating stage' in the evolution of a research programme. Lakatos characterizes this as a stage in which the programme ceases to lead to new discoveries, and ad hoc hypotheses accrue to it that do little to increase its scope. Kuhn writes that 'I cannot myself tell the difference between what he has to say about this important stage and what I have said about the role of crisis in scientific development' (1970: 139).

But Lakatos insists that it is their differing accounts of what ensues from degeneration or crisis that distinguishes his model from Kuhn's. For Kuhn, the shift from one scientific 'paradigm' to another reflects political and cultural factors; for Lakatos, theory development through 'research programmes' is a rational process. Scientists make a rational determination of when to give up on a particular theory. When it stops incorporating new facts it can be considered to be degenerating and, thus, abandoned for a theory that offers an increase of empirical content.

Lakatos was hostile to what he saw as the subjectivist aspects and implications of Kuhn's model and, in particular, the prominent role that Kuhn ascribes to subjectivity in the choice among competing theories and in the shift from one paradigm to another. However, Lakatos does not entirely escape his own critique: subjectivity appears to play a prominent role in his own model, as well.

According to Lakatos, a theory is to be judged on the basis of whether it is able to put forward 'novel facts', some degree of excess empirical content over its predecessor. But progress cannot always be judged in these terms, because not all empirical facts are of equal significance. Novel facts can be either trivial or important, and which they are depends on the values and normative concerns of the individuals making the assessment. It is up to the individual scientist to judge whether a programme that is not generating progressive problem-shifts is merely experiencing a temporary lull or has begun to degenerate; and this, Lakatos says can, in any event, only be known retrospectively. Thus a scientist's 'theory choice' is as much a 'subjective' judgement as the decision to change paradigms is in Kuhn's account.

In comparing these supposedly contrasting views of the nature of the scientific enterprise, what becomes apparent is that, irrespective of whether research employs a positivist or interpretivist approach, or addresses empirical or normative questions, the researcher's subjective valuations are likely to enter into the research process. No approach or type of research ensures a value-free process of inquiry; none can free researchers from the need to be explicit and self-critical concerning their own underlying assumptions and values.

Conclusions

Where does this consideration of the debate about the fact/value distinction leave us?

We began our discussion in this chapter by considering how this distinction has been institutionalized in political research in the division between empirical and normative research and theory. As John Gerring and Joshua Yesnowitz argue, the separation between normative theory and empirical analysis generates problems for both sides of the divide (2006: 104). Empirical study of social phenomena 'is meaningless if it has no normative import; and it is misleading if its normative content is present, but ambiguous'—if we don't know how it matters. Normative arguments that propose or justify one value system over another will lack relevance if they make no attempt to relate to the facts of actual practice or public life. In sum, good social science 'is both empirically grounded and relevant to human concerns' (2006: 133). Normative theorizing 'must deal in facts' and empirical work 'must deal in values' (2006: 108). While we must be sensitive to the difference between normative and empirical questions and statements, we must also recognize that they are not independent of each other, and that there are costs in keeping them separate.

The chapter then explored various ways that 'values' intrude into social-scientific inquiry. It began by considering whether, how, and with what effect, a researcher's own values intrude into research. It then focused on how the act of observation itself intrudes on the object of study. Finally, it took up the question of whether and how research is shaped by shared social practices and is inseparable from them.

To explore this last issue, we considered two different views concerning the role of social factors and practices in scientific inquiry. Thomas Kuhn claims that scientific investigation often reflects relations of power, professional ambition, and wider political and cultural factors; that scientists often accept a theory based, not on observable evidence, but on political and cultural factors. Imre Lakatos emphatically rejects this view of established scientific practice. He argues that scientific inquiry is rational and leads to progressive theory development.

These differing views are really the product of different concerns. Kuhn is concerned with the question of how science advances *in practice*; Lakatos is chiefly concerned with a normative question: how *should* science advance? Kuhn attempts to produce a factual account of science. Critics are right to be concerned with its normative implications. But rather than shooting the messenger, we would be better off treating the message which emerges from Kuhn's account of the history of science as a cautionary tale about how subjectivity and paradigmatic conformity can intrude into research. Lakatos' elaboration of a 'methodology of scientific research programmes' is less a model of how science is actually done, than a proposal for how it *should* be done. It contributes to our thinking about what procedures would allow for a more consistently objective assessment of research findings and pursuit of theory development.

Ernest Nagel has argued that 'steps can be taken to identify a value bias when it occurs, and to minimize if not to eliminate completely its perturbing effects' (1994: 573). To do this we need to proceed with an awareness of the various ways that values intrude into the research process. With this awareness, we can then be in a position to pursue the ideal that Lakatos elaborates: a process of inquiry that, at every step, is both critical and rational.

Questions

1. Is political research value-free?
2. To what extent, and in what ways, do 'values' present problems for political analysis?
3. To what extent, and with what effect, do a researcher's own values intrude into research?
4. To what extent do the values associated with a set of shared social practices intrude into the search for knowledge?
5. According to Thomas Kuhn, what is scientific progress? How does scientific knowledge develop?
6. What is a paradigm? Do paradigms shape research in the social sciences? How?
7. How is Kuhn's conception of a 'scientific revolution' similar to a political revolution?
8. What is Lakatos' notion of a research programme? What aspects of Kuhn's model does it correspond to and differ from?
9. Is the notion of 'normal science' applicable to the study of politics?

Guide to Further Reading

Ball, T. (1976), 'From Paradigms to Research Programs: Toward a Post-Kuhnian Political Science', *American Journal of Political Science* 20(1) (February): 151–77.

Critical Review; Special Issue: Rational Choice Theory and Politics 9(1–2) (1995).

This issue contains a wide-ranging symposium on the criticisms launched against rational choice theory by Donald P. Green and Ian Shapiro, in *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science* (New Haven: Yale University Press, 1994). At least six of the contributions, both pro and con, based themselves in part on their interpretations of Lakatos' theory of research programmes.

Fleck, L. (1979), *Genesis and Development of a Fact* (Chicago: University of Chicago Press).
A fascinating study of the genesis of a theory in medical science that attempts to identify the process by which an idea achieves the status of a fact. Fleck shows that, among other things, a fact must be plausible, and that its plausibility is rooted in a given era. A fact must also be 'suitable for publication', i.e. stylistically relevant within a given culture.

Gerring, J. and J. Yesnowitz (2006), ' "A Normative Turn in Political Science?",' *Polity* 38(1) (January): 101–33.
The authors provide arguments concerning how and why a more normatively orientated study of politics can contribute to theoretical development, empirical inquiry, and disciplinary unity.

Lakatos, I. and A. Musgrave (eds) (1970), *Criticism and the Growth of Knowledge* (Cambridge: Cambridge University Press).
This volume arose out of a symposium on Kuhn's work held in London in 1965. The book begins with Kuhn's statement of his position, followed by seven essays offering criticism and analysis, and finally by Kuhn's reply.

Martin, M. and L. C. McIntyre (eds) (1994), *Readings in the Philosophy of Social Science* (Cambridge, MA: MIT Press).
This volume brings together a collection of important texts on objectivity and values (Part VII). Of particular interest, is Charles Taylor's contribution, 'Neutrality in Political Science' (pp. 547–70). In this influential article, the philosopher Charles Taylor argues that value-neutrality in the social sciences, and the possibility of separating facts and values, is a myth: it is a myth that researchers consider the facts of something and, on the basis of the facts, move to a judgement or valuation of them.

Riley, G. (ed.) (1974), *Values, Objectivity, and the Social Sciences* (Reading, MA: Addison Wesley Longman Publishing Co.).
A collection of articles on research objectivity, values, and partisanship by prominent social scientists.

Science, Technology & Human Values 36 (3) (May 2011). Special issue on 'Post-Normal Science'.

Vasquez, J. A., et al. (1997), Forum on 'The Realist Paradigm and Degenerative vs Progressive Research Programs', *American Political Science Review* 91 (December): 899–934.
Using Imre Lakatos' notion of a 'scientific research programme' (discussed in this chapter), the Forum presented in this issue of the *American Political Science Review* explores the question of whether the 'auxiliary hypotheses' and emendations which various theorists have offered to the Realist Paradigm of International Relations represent the progressive development, or the degeneration, of the paradigm.

References

Ball, T. (1976), 'From Paradigms to Research Programs: Toward a Post-Kuhnian Political Science', *American Journal of Political Science* 20(1) (February): 151–77.

Bernstein, H. R. (1981), 'Marxist Historiography and the Methodology of Research Programs', *History and Theory* 20(4) (December): 424–49.

Bird, A. (2009), 'Thomas Kuhn', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2009 Edition), http://plato.stanford.edu/archives/fall2009/entries/thomas-kuhn/.

Bodenheimer, S. (1971), *The Ideology of Developmentalism: The American-Paradigm-Surrogate for Latin American Studies* (Beverly Hills, CA: Sage).

Cochran, M. (1999), *Normative Theory in International Relations* (Cambridge: Cambridge University Press).

Cox, R. (1981), 'Social Forces, States and World Orders: Beyond International Relations Theory', *Millennium: Journal of International Studies* 10(2): 126–55.
Diamond, S. (1992), *Compromised Campus* (New York: Oxford University Press).
Foster-Carter, A. (1976), 'From Rostow to Gunder Frank: Conflicting Paradigms in the Analysis of Underdevelopment', *World Development* 4(3) (March): 167–80.
Frankel, C. (1969), *High on Foggy Bottom: An Outsider's Inside View of the Government* (New York: Harper & Row).
Funtowicz, S. O. and J. R. Ravetz (1992), 'Three Types of Risk Assessment and the Emergence of Post Normal Science', in S. Krimsky and D. Golding (eds), *Social Theories of Risk* (Westport, CT: Praeger), 251–74.
Funtowicz, S. O. and J. R. Ravetz (1993), 'Science for the Post-Normal Age', *Futures* 25(7): 739–55.
Geddes, B. (2003), *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics* (Ann Arbor: University of Michigan Press).
Gendzier, I. (1985), *Managing Social Change: Social Scientists and the Third World* (Boulder, CO: Westview).
Gerring, J. (2001), *Social Science Methodology: A Criterial Framework* (Cambridge: Cambridge University Press).
Gerring, J. and J. Yesnowitz (2006), 'A Normative Turn in Political Science?' *Polity* 38(1) (January): 101–33.
Gutting, G. (1984), 'Paradigms and Hermeneutics: A Dialogue on Kuhn, Rorty, and the Social Sciences', *American Philosophical Quarterly* 21(1) (January): 1–15.
Higgott, R. (1983), *Political Development Theory: The Contemporary Debate* (Beckenham: Croom Helm).
Jastrow, J. (1899), 'The Mind's Eye', *Popular Science Monthly* 54: 299–312. (From Wikipedia, http://en.wikipedia.org/wiki/file:Duck-Rabbit_illusion.jpg)
Jervis, R. (2002), 'Politics, Political Science, and Specialization, PS', *Political Science and Politics* 35(2) (June): 187–9.
Kuhn, T. (1962), *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press).
Kuhn, T. (1970), 'Notes on Lakatos', *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 137–46.
Lakatos, I. (1970), 'Falsification and the Methodology of Scientific Research Programs', in I. Lakatos and A. Musgrave (eds), *Criticism and the Growth of Knowledge* (Cambridge: Cambridge University Press), 91–196.
Martin, M. (1994), 'The Philosophical Importance of the Rosenthal Effect', in M. Martin and L. C. McIntyre (eds), *Readings in the Philosophy of Social Science* (Cambridge, MA: MIT Press), 585–96.
Mason, M. (1997), *Development and Disorder: A History of the Third World since 1945* (Hanover, NE: University Press of New England).
Morgan, P. M. (1994), *Theories and Approaches to International Politics: What Are We to Think?*, 4th edition (New Brunswick, NJ: Transaction Publishers).
Nagel, E. (1994), 'The Value-Oriented Bias of Social Inquiry', in M. Martin and L. C. McIntyre (eds), *Readings in the Philosophy of Social Science* (Cambridge, MA: MIT Press), 571–84.
Pavlich, G. (1988), 'Re-Evaluating Modernisation and Dependency in Lesotho', *The Journal of Modern African Studies* 26(4) (December): 591–605.
Popper, K. (1959), *The Logic of Scientific Discovery* (London: Hutchinson of London).
Schmidt, B. (2002), 'On the History and Historiography of International Relations', in W. Carlnaes, T. Risse, and B. Simmons (eds), *Handbook of International Relations* (London: Sage), 3–22.
Simpson, C. (ed.) (1999), *Universities and Empire* (New York: New Press).
Valenzuela, A. and S. Valenzuela (1979), 'Modernization and Dependence: Alternative Perspectives in the Study of Latin American Development', in J. Villamil (ed.), *Transnational Capitalism and National Development* (Sussex: Harvester Press), 31–65.
Weber, M. (1948), *The Methodology of the Social Sciences* (New York: The Free Press).
Weber, M. (1994), 'Objectivity in Social Science and Social Policy', in M. Martin and L. C. McIntyre (eds), *Readings in the Philosophy of Social Science* (Cambridge, MA: MIT Press), 535–46.

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

Part 2

How to Do Research: An Overview

4 Asking Questions: How to Find and Formulate Research Questions 93

5 Finding Answers: Theories and How to Apply Them 123

6 Research Design 159

7 What is Data? 186

Asking Questions: How to Find and Formulate Research Questions

Chapter Summary

In this chapter we move from issues in the philosophy of social science and begin our consideration of the research process. We start by considering *step one* of this process: the formulation of a well-crafted research question. As we shall explain, a research question not only initiates the research process, it is crucial to every other step along the way. This chapter discusses why your research should begin with a research question, and how it structures the research process; the difference between a topic or general question, on the one hand, and a focused research question, on the other; where to find and how to formulate research questions, the various types of questions scholars ask, and the role of the 'literature review' in providing a rationale for research questions. In discussing these issues, the chapter addresses a number of questions, including the following:

- What is a research question?
- Why begin political inquiry with research questions?
- What kinds of questions stimulate good research?
- How can you develop a research question from the existing literature relating to your area of interest?
- What is the role of the 'literature review' in a research project?
- What form does a good research question take?

Introduction

> Questions are the engines of intellect, the cerebral machines which convert energy to motion, and curiosity to controlled inquiry.
>
> (Fischer 1970: 3)

Most political research originates from some general question or problem that arises either from the events, issues, or processes we observe in the world around us, or from the theories and frameworks that our field has developed in order to understand them. But the

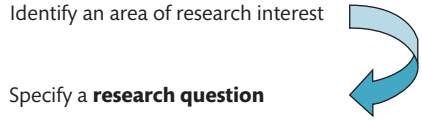


Figure 4.1 Step one

research *process* only begins with the conversion of this general question or problem into a well-formulated, clearly focused, research question. As we shall explain, there is an important difference between a *general* question, topic, or idea that you may wish to investigate and a *research* question. What the formulation of a research question enables you to do is to channel your interest, intellectual energy, and curiosity into the pursuit of structured, systematic inquiry. This first step, in which you move from identifying a topic to defining a research question (illustrated in Figure 4.1), can be challenging, frustrating, and time-consuming. But once it's accomplished, you are on your way!

Consequently, this chapter is devoted to this important fist step. It discusses:

- the requirements of a research question
- where research questions come from
- how to *survey* the literature in a way that provides you with an overview of the topic that you are interested in
- how to write a '*literature review*' which provides a rationale for your research question (and for your answer to it)
- the different types of research questions scholars in our field ask
- how to formulate a research question so that it addresses an issue of significance both in the real world and for a broad community of scholars and political researchers.

There are three misconceptions about research questions that we should clear up at the outset.

The first has to do with the role and function of a research question. When we say that good research *begins* with a well-formulated research question, this might be interpreted as meaning that a research question is important as a means of *initiating* the research process. But a research question plays a continuous role throughout the research process. It establishes the basis for every decision a researcher makes in every phase of a project. So, getting the research question fully in focus not only gets you off to a good start—it informs every other part of the research process.

A second misconception concerns the task of formulating a research question. This may sound like a relatively straightforward task—one that is easily and quickly accomplished; and maybe you are wondering why we need to devote an entire chapter to discussing how to go about it. However, formulating a good research question is not as easy as it might first appear. It often demands considerable time and thought and, for many researchers, may be the most difficult part of a research project. But however much time it might take to accomplish, once you have formulated a good research question, the rest of the research process becomes far easier and clearer, and moves along much faster.

A third misconception about research questions has to do with their relevance to different types of research. Some people may think that the formulation of a specific research question is relevant only to quantitative research; that qualitative research involves a different, more open-ended process and so doesn't require the framing of a research question. But, irrespective of whether you are pursuing qualitative or quantitative research, there are good reasons why your research should be based on and shaped by a research question. Below, we list six reasons why formulating a research question should be step one of the research process:

1. It forces you to get clear about what you want to know.
2. By requiring that it address an issue of significance in the real world, it ensures that your answer, if systematically developed, will make it possible for your research to contribute to knowledge about important issues.
3. By requiring that it addresses an issue of significance to the field of politics, it ensures that your answer in some way contributes to, rather than merely reproduces, existing knowledge.
4. It organizes, and is the basis for, everything else that follows.
5. It focuses and narrows your search for information.
6. It enables you to decide what to include and what to exclude from your research.

Consider the steps involved in conducting research outlined in Box 4.1: these show that every component of the process flows from and is directly connected to a research question. We will be discussing each of these components over the course of this chapter and Chapters 5 and 6.

But before beginning this discussion, we ask you to note two points we wish to emphasize about the outline of research components presented in Box 4.1, as well as the discussion throughout this chapter.

First, this outline is meant to provide a *general template* for formulating a research question and structuring research. The design of *your* research will ultimately be shaped by your specific topic and the aims of your project, and by the guidelines that your thesis/dissertation supervisor or course tutor requires that you follow. 'Your answer' part of this outline suggests a focus on hypothesis generation and theory testing, an approach not equally relevant to all types of research. It might be argued, too, that this is an approach not accepted in the different forms of knowledge (positivism/interpretivism) discussed in Chapter 2. But as Chapter 2 pointed out, though positivist and interpretivist researchers tend to be guided by different assumptions, they also tend to share a common set of research practices founded in the hypothetico-deductive method. Some researchers 'assess whether the information they have gathered fits with the interpretation they have posited', and others 'consider the fit of competing interpretations with the facts they have gathered', but 'in either case they are practicing the hypothetico-deductive method' (Pollins 2007: 100). Consequently, they generally follow the same methodological conventions. But, again, Box 4.1 serves as a starting point and set of considerations to be revised or elaborated consistent with the question and aims that animate your own project.

Second, the research process that Box 4.1 shows is one that is structured in an idealized linear fashion. The real process is different. For instance, though researchers may start with 'some operational hunches about what they expect to find', they generally do not '"frame

BOX 4.1 Outline of Research Components

1. **The question**
   - A. What do you want to know? What is the *central question/problem/issue/puzzle*?
   - B. Why do you think it is worth doing/knowing? What is *the rationale for pursuing research on this question*?
2. **The literature**
   - A. Who else has asked and answered your question? What (range of) answers to this question are found in *the existing literature relating to it*?
   - B. What are the *positive elements* in the current literature? What in the literature can you highlight, underline, expand, extend, improve, build upon, continue?
   - C. What needs to be done? Delineate the crucial aspects of the problem requiring investigation. *What needs to be done in order to provide a better answer to the question than currently exists*?
3. **Your answer**
   - A. *Theoretical framework*. What are the *theoretical elements and guiding assumptions of the study*?
     1. What factors or variables of the problem must be investigated in order to answer your central question?
     2. What is/are *your hypothesis/es* (how are these factors linked)?
        - a. What is the source of your hypothesis/es? What in theory would lead us to expect the relationship(s) you assert?
        - b. How would you demonstrate the relationships stated by the hypothesis/es?
     3. What is *the spatial/temporal domain of the study*? What is the rationale for defining this domain for the study?
   - B. *Data and sources*
     1. What are *the data relevant to demonstrating the relationships you hypothesize*?
     2. What *sources* are there *for these data*?

hypotheses" in any formal sense before they start to work' (Shively 1989: 25). Good researchers generally remain engaged in learning and creative re-thinking all through the research process. As Chapter 5 will discuss, our aim in presenting the outline shown in Box 4.1 is to offer some clear and obvious procedures to get you started. As you gain experience, you will develop your own process of work and way of fulfilling the requirements of good research.

Irrespective of the particular way you choose to structure your research, a carefully formulated research question will ensure that your research has a clear purpose. If you are not clear what question or questions your research is designed to address, it will be difficult to produce research that is focused and, ultimately, interesting to others. More specifically, a research question ensures that your research focuses on and is clearly linked to the realization of a specific aim that contributes to our knowledge of some significant problem or issue; and that your research builds on existing knowledge. A properly formulated research question also provides, as a general question or topic does not, the basis for a *logic of inquiry* to guide decisions about the design and methods of the research—decisions about what information you should include or exclude, what data you will need to collect, what to observe, or what to ask your interviewees. Once you have formulated a research question, the research task becomes specific so that, rather than trying to gather all the information you can find on a topic, you direct your attention to just the information that addresses the question and helps you to develop an argument.

Research questions: what are they? Where do they come from?

What is a research question? A research question is one that (1) has significance for a topic or issue relating to the subject matter of our field; (2) is researchable (it can be answered through conducting research); and (3) has not yet been answered definitively.

All three of these requirements reflect a basic attribute and requirement of social science research: that it addresses a topic or issue relating to the subject matter of a field of study in a way that contributes to our knowledge of that topic. The first requirement compels you to relate a specific topic that interests you to a broader issue or area of inquiry within the field of politics ([1B] in Box 4.1). The requirement that a question be researchable means that it must be formulated in a way that makes it possible for you to provide an answer to it through conducting research ([3] in Box 4.1). To be researchable it will have to be sufficiently focused to permit an answer that can in some way be tested or demonstrated, and free of errors of reasoning that might generate false answers or conclusions. Research is expected to contribute empirically, conceptually, or theoretically to our knowledge of something. Consequently, the third requirement—that a question must be one which has not been definitively answered—requires that you explicitly locate your research within the framework of an existing social scientific literature ([2] in Box 4.1). Doing this ensures that you will proceed with an awareness of the 'state of the art' in terms of existing knowledge on the topic or issue that interests you, and so minimize the chance of duplicating what has already been done. In order to claim that your research question 'has not yet been answered definitively' you must show that your question has not been addressed; or that existing research has failed to resolve debates about your question; or that there is still more that can be done to improve the answers that have so far been offered by, for instance, providing a more systematic, thorough answer, or by considering the question in light of different cases or perspectives that might provide further illumination.

We will discuss each of these requirements a bit further on when we consider how to formulate a good research question. First, let's consider how to go about finding a research question.

Finding research questions

Research *topics* in our field originate from our observation of real-world problems and events such as conflicts, elections, or revolutions. But an important source for finding research *questions* is the literature in the field of politics as well as in social science literatures relevant to the subject matter of our field. Please note that, while a *topic* is usually a statement, a *research question* is a question: it is a sentence with a *question mark* at the end of it.

Before we discuss how to use the literature to find a research question, there are two caveats worth noting.

First, though the literature in our field is an important source for research questions, you should not approach the literature with too accepting a frame of mind. As Barbara Geddes points out, in our field:

> [t]he literature on some subjects contains only a few arguments generally accepted as true; many controversies in which the hypotheses on both sides lack both clarity and strong empirical support; and large amounts of opinion and conjecture, unsupported by systematic evidence but nevertheless often referred to as theory.
>
> (2003: 29)

Students are frequently advised to 'look for holes in the literature' in order to find research questions. But Geddes' characterization of some of the literature in our field suggests that the literature is important to the formulation of a research question, *not* because it suggests gaps that need to be filled, but because it stimulates 'indignation, annoyance, and irritation' (Geddes 2003: 29–30). It should be noted, too, that our understanding of the political world usually increases, not through finding and filling gaps in existing knowledge—as Patrick Dunleavy points out, there are often good reasons why gaps remain unfilled (2003: 274)—but by well-crafted research that produces incremental extensions and revisions of it.

In any case, it is worth emphasizing that you should approach the literature on any subject analytically. To be 'analytical' is to ask questions: to subject something to questions in order to discover its meaning or essential features. So our advice is to approach the literature with a questioning frame of mind: one that combines respect for the efforts and achievements of the scholars who produced it, with a healthy degree of scepticism.

Our second caveat about using the literature as a source for research questions is that you should avoid getting drawn into questions that are politically insignificant. What we want to alert you to here is that some portion of the literature on your topic may be devoted to questions that, having become the focus of dispute among academic researchers, have lost connection with the real-world problems that they were originally meant to address. Research is expected to contribute to our knowledge of a real-world problem or issue, and we can address a number of different kinds of research questions that enable us to do this. In addition to asking empirical questions about real-world problems, we can also ask conceptual, theoretical, and methodological questions which address shortcomings in the theories and tools our field has developed in order to understand these problems. Since we do not yet have fully effective tools for investigating the many important problems that researchers want to address, research that makes the tools we use to understand a given problem more effective also contributes to existing knowledge. Our point in raising this caveat is that research which focuses 'too much on making a contribution to a scholarly literature without some attention to topics that have real-world importance' runs the risk of 'descending to politically insignificant questions' (King, Keohane, and Verba 1994: 17). The best research manages 'to be directly relevant to solving real-world problems and to furthering the goals of a specific scientific literature' (King, Keohane, and Verba 1994: 18). Box 4.2 sums up this view.

With these caveats duly noted, it should be emphasized that reading the literature on a specific topic of interest to you is an important means of gaining knowledge both of existing research and current debates, and of suggesting questions about the topic.

Any issue relating to the political world is likely to be the subject of debate. Experts are likely to produce a variety of conflicting views about how the issue should be to defined; where, when, and why it originated; how it developed or evolved; what its impact is; what its likely future trajectory will be; and what decision-makers or specialists can or should do with respect to it. If you plan to pursue research on an issue, you will need to sort through

BOX 4.2 What Do We Mean When We Say that a Research Project Should Be 'Significant'?

FIRST: *a research project should pose a question that is 'important' in the real world*. The topic should be consequential for political, social, or economic life, for understanding something that significantly affects people's lives, or for understanding and predicting events that might be harmful or beneficial.
SECOND: *a research project should make a specific contribution to an identifiable scholarly literature* by increasing our collective ability to construct verified scientific explanations of some aspect of the world.

(King, Keohane, and Verba 1994: 15)

the various opinions and judgements about it and come to your own conclusions. But first you will need to get a feel for the debates—for the different positions, and the current 'state of play'. You will then be able to figure out how to make visible some dimension of the issue that brings added strength to one or another of the positions represented in that debate, produce a synthesis of their elements, or show how the debate itself is based on erroneous assumptions or logic.

A good way to go about doing this is to survey the literature on the topic that interests you. We are not referring here to a literature *review*, which is something you write *after you have formulated your research question* (we will be discussing this later in this chapter). A *survey* of literature on your topic helps you to focus your interest and your search for a question on a narrower, more specific, aspect of the topic.

Let's say, for example, that you are interested in problems of democracy, and that you want to learn more about it and make it the focus of a research project. This is a topic of great importance and interest to researchers throughout the social sciences, and the literature relating to it is immense. So, how can you convert your general interest in this large subject into a well-formulated, clearly focused research question—one that will engage your interest, channel your energies, and enable you to contribute to current discussions and debates about some aspect of the subject?

What we recommend is that you *survey* the literature about democracy in order to find out what questions other researchers are asking about the topic and what key debates and issues their research addresses. Look at the list of questions in Box 4.3. These represent only a small sample of the questions that researchers ask about democracy. But each question alerts you to a different avenue of research and directs your attention to a more specific literature and set of research questions. You can then select those you think you might be interested in pursuing and compile a bibliography of previous research and writing on them. This, then, is the aim of the survey: to compile a starting bibliography on a narrower aspect of the general topic that interests you. With this done, you read the titles on your bibliography (and the abstracts or summaries of the entries, as well); and this will either inspire ideas about how to extend, refine, or critique this literature, or direct your attention to other literatures and avenues of research.

Here is a simple strategy you can use to get this process going.

1. Find, through an online search, a book whose title seems to promise either an introduction to an interesting facet of your topic or a good general survey of the topic, and turn to

BOX 4.3 Twenty Questions about Democracy

DEMocracy

1. What are the problems of democratization in newly decolonized countries?
2. What are the implications for democracy of globalization?
3. Do democratic states have more in common with each other than with undemocratic states?
4. Are there different models of democracy in different cultural contexts?
5. What are the different institutional manifestations of democracy?
6. Is democracy always liberal?
7. Can democracy be imposed upon countries from outside or must it draw upon the established practices of countries *even if* these practices are not obviously democratic?
8. To what extent, and on what conditions, might democracy be imposed (hence undemocratically) by a foreign power?
9. Are interest groups a threat to democracy?
10. Is the concentration of media ownership a threat to democracy?
11. What socio-economic factors are positively associated with democracy?
12. How does electoral system design affect ethnic conflict regulation in ethnically divided societies?
13. How essential are political parties to representative democracy?
14. What is the relationship between nationalism and democracy?
15. What is a developmental state? Is democracy relevant to its conceptualization?
16. Can ethnic/religious/linguistic politics and democracy be positively related with each other?
17. Is democracy the best political system for promoting human rights in developing countries?
18. Can democracy promote interstate conflict resolution?
19. Does democracy require an explicit understanding and justification of certain core values?
20. What are the institutional implications of deliberative democracy?

its bibliography or list of references. The reference section of one well-chosen source will contain pages filled with titles that indicate questions and issues people are writing about. These will stimulate ideas of how you might narrow your focus. Moreover, when you go to the library to get the book, you will discover on the library shelf where the book is kept that it is located in a neighbourhood populated by many other books on the same topic. You can then do a 'neighbourhood search' (peruse the titles of the books on the surrounding shelves) to get a sense of the variety of facets of and entry points into your topic.

In addition, or alternatively, you will want to search the JSTOR database: http://www.jstor.org/. JSTOR contains the full text of more than 220 peer-reviewed political science journals. Enter keywords in the full-text search. Make sure that the 'Political Science Journals' box is checked; and, if your topic involves specific geographical areas (e.g. the Middle East, Latin America) or issues and factors central to other disciplines (e.g. Economics, Anthropology, Psychology), you may want to select other journal categories as well. Peruse the resulting list of articles for three or four that seem particularly relevant, and read their abstracts, introductions, and conclusions. You will also want to search the Web of Science (http://webofknowledge.com/). This source includes five databases covering virtually all fields of knowledge. Two of the most useful are the *Social Sciences Citation Index* (*SSCI*) and the *Arts & Humanities Citation Index (A&HCI)*. When you find a scholarly article relevant to your particular research interest, you can also locate all subsequent

journal articles that cite it. Also check Google Scholar (http://scholar.google.com/). *Always be sure to retain full bibliographic references (author, date, title, journal, volume, pages) for the articles of interest you find.*

2. A particularly useful source to get you started is a survey or 'state of knowledge'-type article. These often provide a breakdown of the literature on a topic according to its various entry points, debates, and perspectives, and provide a review of the evolution of thinking on the topic up to and including the most recent and influential works. See if you can spot a recent article on your topic whose title conforms to something along the lines of the following: 'The Study of Gender and Politics: The State of the Art' or 'Parliamentary versus Presidential Systems: A Survey of the Literature'. Relevant course syllabi or reading lists, which may be accessible online, can also provide this sort of introduction to the literature on a topic.

3. Pursue the line of investigation that seems most interesting to you by getting one of the relevant books or articles you find in a bibliography (or on a course reading list), and then turn to the back of *that* book or article to get additional titles (and also to get 'search terms' that you can put into a search engine).

Bibliography or reference management software packages allow you to save and store the bibliographic details of the references you find (articles, books, conference papers, websites, etc.). With these packages you have not only stored the bibliographic details of the references you think you may want to use, you also have a personal database of references that you can search and sort; and you can connect between your package and Word to create a customized bibliography in your preferred referencing style at the end of your document. If you change your mind about what referencing style you want or need to use, you can quickly and easily reformat the references. Some of the more popular packages are Endnote, RefWorks, Zotero, Mendeley, and CiteULike.

4. Once you have a bibliography of promising books and articles on a particular line of investigation, read the literature you have compiled and think, critically and creatively, about its insights and how to extend them, or its weaknesses and flaws and how to overcome and correct them.

Box 4.4 provides a snapshot of how a survey of the literature on a topic reveals numerous different dimensions or facets of a specific topic.

Institutional libraries offer abundant support to assist you, including online and paper-based guidance on searching databases and finding research literature, web searching, how to reference (including citing electronic resources) and managing references, and new developments in search tools and websites.

Define your topic in terms of 'keywords' that you can use to search various information sources. Think about alternative meanings and contexts for your keywords. Also think of ways of combining words to ensure that only the meaning you want is retrieved. And consider alternative spellings of the words you use. For instance, American and British English use different spellings for many words. So you may want to use a truncation that will capture both spellings (such as behavio*, which will retrieve behavior, behaviour, behavioral, and behavioural).

Look at ways to link your keywords. Do you want records which contain all of your keywords, or are some of them alternatives? Do you want to relate two words so that they are near each other? Can you focus your search by excluding words which might appear with

BOX 4.4 Surveying the Literature to Narrow Your Search for a Research Question

Step one

1. Identify a research area: *conflict*.
2. Identify a case, particular space of time, or geographic area in which you might want to investigate it: *Mexico*.
3. In Google Scholar type '*Conflict Mexico*'. Scan the titles of the listed articles. This identifies questions that scholars are asking about these conflicts, and suggests questions that they either fail to address or address insufficiently, including the following:
   a. What are the motivations behind the conflicts?
   b. What are the main economic determinants?
   c. What are their relationships with political and social issues?
   d. What type of conflicts persist in the country?
   e. What are the root causes of the conflicts?
   f. What are some of the politico-economic losses and gains?
   g. What have been the impacts on human development and poverty?
   h. How does the analysis of the conflicts relate to past and current theoretical discourse and debate?
   i. Which and where are the inflexion points of the conflicts?
   j. Have the causes and relationships changed over time? What dimensions have influenced them?
   k. What is the intensity of the conflicts?
   l. How does (do) the conflict(s) in Chiapas compare to other cases in Mexico? Are there any linkages?
   m. What issues have been addressed? Which ones remain unresolved?
   n. What politico-economic tools are needed to mitigate and overcome the conflicts?
   o. What are the lessons learned for conflict prevention and conflict resolution?
4. Select an aspect or dimension of the research area and compile a bibliography on it: *the relation of environmental factors to persistent internal conflicts in south-eastern Mexico (the states of Chiapas, Guerrero, and Oaxaca)*.

Step two

5. Begin to assemble a bibliography on this research area by scanning one article, below: 'Environmental Scarcity and Violent Conflict: The Case of Chiapas, Mexico' (Philip N. Howard and Thomas Homer-Dixon), Occasional Paper, Project on Environment, Population and Security, Washington, DC: American Association for the Advancement of Science and the University of Toronto, January 1996. Pick out the particularly interesting titles cited in this article and from among the thirty-two titles that Google Scholar shows as having cited this one, including this particularly useful looking one: Nils Petter Gleditsch, 'Armed Conflict and the Environment: A Critique of the Literature', *Journal of Peace Research*, Vol. 35, No. 3, Special Issue on Environmental Conflict (May, 1998), pp. 381–400.
6. Collect, and then read, the literature you have compiled on that specific aspect or dimension of the research area, and think of ways to extend or refine it.

other words you are searching for but in a different context? How you do this and what options are available are not the same with every database, so you will need to use the on-screen links and help system to familiarize yourself with how each of them works.

Set limitations on your search as, for instance, by publication date (how far back you want to search), range (the types of publication or documentation you wish to include), geographical or time scope, or country of publication. Be sure to keep a complete and accurate record of your searches. A detailed record of everything useful you find will enable you to provide an accurate bibliography at the end of your project.

As part of a strategy to narrow your search for a question, you might do an online search of fellowship and grant programmes that provide support for research in our field. There are various agencies and funding bodies that offer awards and grants to support research on topics related to the subject matter of our field. And often they issue calls for applications or proposals. These calls for proposals typically focus on a broad topic and include a breakdown of that topic into a set of narrower issues or questions relating to it. Box 4.5 shows a sample of a call for proposals in the field of politics. Government agencies also publish requests for proposals. These are likely to describe some problem that the agency would like researchers to address, and the approach they would like them to take in investigating it. We would not encourage you to actually respond to these calls while working towards a degree. But they might serve to spark your interest in a question and start you on the path that leads you to a research question.

BOX 4.5 Calls for Proposals as a Source for Research Questions

Global uncertainties: security for all in a changing world programme, under the auspices of Research Councils UK

ESRC/AHRC fellowships on ideas and beliefs

The ESRC and AHRC invite full applications from researchers from across the social sciences, arts, and humanities for a number of high-profile research fellowships, as part of the Research Council UK (RCUK) Global Uncertainties: Security for All in a Changing World programme.

The cross-Council programme focuses on the nature and interactions of five global issues: conflict, crime, environmental degradation, poverty, and terrorism, and their implications for various concepts and contexts of security and insecurity. Within this framework, this fellowship call focuses specifically on how ideas and beliefs of individuals, communities, and nation states relate to these five global phenomena. Fellowship applications under this call must address one or more of the following key research areas:

a. How do individuals and communities develop their ideas and beliefs about security and insecurity?
b. Why do some ideas and beliefs lead to conflict, violence, or criminal activity? What lessons can we learn from (a) above that provide the basis for countering those ideas and beliefs that reinforce conflict, violence, and crime?
c. How do issues around the cycle of knowledge production and use interact with the creation, management, and resolution of insecurities?
d. How are risks and threats communicated, constructed, represented, and received by key actors and communities, using different media and cultural forms for different audiences, including the use of language, images, and symbolism?
e. Is there an acceptable balance between national security needs and the protection of civil liberties and human rights? If so, can one be secured? And how do we balance local needs against global responsibilities within a security context?
f. How should institutions with responsibility for different aspects of a broad security agenda, including security forces themselves, evolve to meet new risks and threats?

Proposals addressing the interaction between these questions/areas and applications which challenge existing policy or practice assumptions and/or consider alternative framing or approaches to addressing these priority research areas will be welcomed. The full specification document can be found in the below link.

Source: http://people.bath.ac.uk/esscjs/Research_Office/GU_Fellowships_Spec.pdf

We have discussed where and how to find a question. This is the best way we can think of to get you started. However, as you develop experience and confidence as a researcher, you will doubtless be able to find questions in other, and perhaps more creative, ways.

In any case, once you have a question, it is likely that there will still be some distance to travel before you have formulated it so that it satisfies the three criteria which define a research question (one that has significance for a topic or issue relating to the subject matter of our field, that is researchable, and that has not yet been answered satisfactorily). We now take up the issue of how to formulate a research question.

How to formulate a research(able) question

Let's begin our consideration of how to formulate a research question by recalling the three requirements of a research question. For the moment, we will focus on the first two requirements; we will be discussing the third requirement a bit further along.

At the beginning of the research process, the requirement that a question be both (1) *significant* (for real-world problems or for academic debates or theories) and (2) *researchable* (answerable through research) can sometimes be a source of frustration. This is because a researcher will often start with an interest *either* in a general subject matter or broad question, *or* in a quite specific case or problem. Those who begin with an interest in a general subject matter sometimes have difficulty coming up with a more specific case or domain in which to investigate it; that is, they have difficulty coming up with a narrower, researchable question. Conversely, those who begin with a narrower issue or problem sometimes have difficulty relating it to a broader, more general area of interest. So we often have in hand a significant question that we don't know how to make researchable, or a researchable question whose significance for the field of politics we have not yet been able to determine or define.

To aid the process of formulating the question in a way that meets both these requirements, we introduce a tool which we call the 'research vase'. By providing a visualization of the research process, the 'research vase' may help you to locate where your question falls in the typical conundrum we have described, and to identify what you need to do in order to make it satisfy both requirements of a research question.

The research vase

Most research projects share the same general structure. We can think of this structure as following the shape of the vase that is pictured in Figure 4.2.

The vase in this picture has an hour-glass shape: it has a broad top and base, and a narrow middle. Figure 4.2 shows how different aspects of the research process can be seen as corresponding to different parts of the vase. The top of the vase represents a large question that engages a broad theoretical perspective or area of inquiry relating to the subject matter of our field. Its narrower middle represents a narrower, more specific question which can be addressed through conducting research: a question which is, in other words, 'researchable'. After passing through this narrow middle, the vase broadens out again. This part of the vase represents the final stage of research: the stage in which the researcher reflects on the

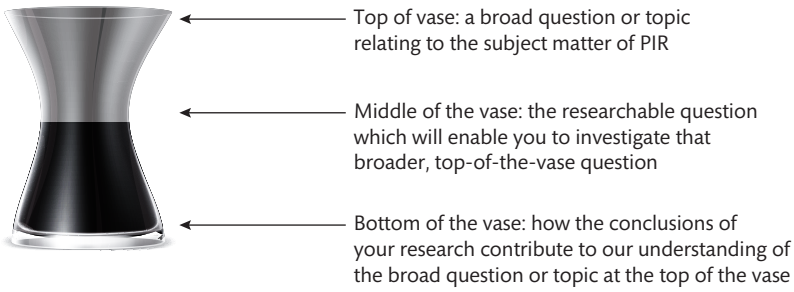


Figure 4.2 The research vase

Source: Kolonko/Shutterstock.com.

conclusions of the research (that was conducted on the 'middle of the vase' question) and the implications of these conclusions for our understanding of the broader, top-of-the-vase question or larger area of inquiry.

The vase, then, enables us to visualize the first two requirements for a research question and how they are related to each other. The *top of the vase* represents the broader question or terrain within which you locate your research, and for which it is significant. When you 'fill in' this part of the research vase you satisfy the first requirement of a research question (that it has significance for a broad community of scholars), and you also establish part of the *rationale* for your question.

If you look again at the components of the research process shown in Box 4.1, you'll see that 'the question' includes two elements: a statement of the question which your research will address; and the *rationale* for pursuing research on it. We establish a rationale or justification for pursuing research on a particular question by showing that it meets two of the requirements of a research question: that it addresses a question that has real-world significance, and that it has not yet been definitively defined. With regard to the first of these two requirements, the rationale that you provide for your question explains *the context* for its consideration. Does it engage a current real-world problem or issue? Does it arise from an issue that is the focus of debate in our field? The rationale also provides a *justification of its importance for study*. Why do you think what the question asks is worth knowing? Why does it matter intellectually, theoretically, or empirically? So, with regard to the first requirement, the rationale for pursuing research on a question rests on a statement of its broader significance and how its answer will contribute to collective knowledge about politics and international relations. But the rationale will also indicate something about the second of these two requirements, by saying something about why and in what respect the question has not yet been definitively answered.

Let's move on to the middle of the vase. This represents the narrower question that will be the focus of your research. Answering this narrower question engages the researcher in direct measurement of something or specific observations of reality in a particular temporal and/or spatial setting. This, in turn, enables the researcher to deaw conclusions which contribute in some way to our knowledge about the question at the top of the vase. In sum, the role of the middle-of-the-vase question is to enable the researcher to address the broader topic or question which is at the top of the vase and which, either as stated or in respect to what it asks, cannot be directly answered. The bottom of the vase represents the

stage in the process of research where the researcher uses the conclusions of the research to reflect back on the broad question or topic at the top of the vase.

We previously noted that it is often the case that, at the beginning of the research process, a researcher will have in mind *either* a (top-of-the-vase) question that is too broad to be researchable, *or* a (middle-of-the-vase) question that is very specific and needs to be related to a broader issue or problem of significance in our field. Those with an interest in a general subject matter will need to find an empirical case or domain—a specific aspect of the problem and/or a setting (a specific country or group)—which will enable them to investigate the subject matter and to derive conclusions relevant to the broader question; those who start with an interest in a particular case will need to identify it as a case, instance, or aspect of a more general issue of interest to scholars pursuing research in the field of politics.

Consider this question: 'Is the US winning its "War on Terror"?' This is a good question: it is significant, the subject of much debate, and one to which we do not yet have a definitive answer. But it is also a question that *in this form* is not possible to answer through research, because it is too broad to be studied directly. It is, in other words a top-of-the-vase question. Consequently, if you wish to pursue research on this question, you would need to express it in a form that is *researchable*. This would require you to find an aspect of the 'War on Terror'—one weapon or policy—that can be the focus of research, and a setting in which to study it. For instance, you might focus on one initiative being pursued in the 'War on Terror': the securitization of Muslim charities. 'Securitization' means the representation of something (in this case, the representation of Muslim charities by governments) as a potential security problem. In the case of Muslim charities, they have been securitized because their finances may support groups classified by governments as connected to illegitimate violence. With this as a focus, you might ask: 'What has been the impact of the securitization of Muslim charities in Egypt since 9/11?' You might hypothesize that 'the securitization of Muslim charities in Egypt since 9/11 has *increased* tensions within the country'. Now you have developed an empirically testable hypothesis which can contribute to answering the broader question with which you began.

Alternatively, you might have begun the research process with an interest in a narrower (middle-of the-vase) question about the impact of new restrictions on Muslim charities in Egypt. Here, you begin with a specific research interest, and will need to 'fill in' the 'top of the vase' by considering what more general phenomenon or issue the research question addresses; how your question might be linked to a broader area of concern or inquiry in our field; what your findings might eventually tell us about more general issues. This, then, might lead you to position your research on Muslim charities in Egypt as part of a broader question concerning whether or not the US is winning its 'War on Terror'.

To continue, for a moment, with the issue of relating a specific question to a broader terrain: you can often identify the broader empirical or theoretical terrain or context relevant to your question by translating individual names into 'kind' names. For instance, an interest in the ASEAN security framework can provide insight into the more general phenomenon of 'regional integration'; a specific question about the north–south conflict in Sudan can engage an aspect of the literature on 'civil wars'. A researcher who wants to learn more about the tribal Pashtuns of south-eastern Afghanistan and north-western Pakistan might relate this research to a bigger question about the survival and continuing relevance of tribes as a form of political organization and authority.

But let's look more closely at what is entailed in the process of 'narrowing down' a broader question. Consider the following question: 'Does globalization undermine the basis of the nation state?' The question concerns the relationship between 'globalization' (concept 1) and 'the basis of a nation state' (concept 2). To make this question researchable, each concept will need to be defined specifically enough to allow us to observe or measure whether, or to what extent, one of them (globalization) is adversely affecting (undermining) the other (the nation state). To do this we can choose one facet of globalization and one of the specific features that form the basis of the nation state, explore the relationship between the two, and then reflect on what the conclusions of this research tell us about the relationship between globalization and the basis of the nation state. We might choose, for instance, to focus on economic integration, which is a key facet of globalization, and on state autonomy, which is a key basis of the nation state. This will enable us to ask a narrower question: 'What impact does economic integration have on state autonomy?' Through further reading, we might discover that researchers measure the integration of a country into the world economy by using an index that combines measures of a country's various transnational links and its involvement in international organizations. Further reading will also enable us to better understand the core functions of the nation state; and we might decide that a state's ability to manage public finances is a critical indicator of the autonomy of a state, inasmuch as a state cannot enjoy autonomy if it must depend on external sources to fund its ongoing operations. We continue to explore existing research and find that researchers have measured the degree of a state's autonomy and whether it is increasing or decreasing by using data on the ratio of domestic to foreign sources of revenue in a state's budget at a given time, and changes in this ratio over time. Our question, then, might be: 'Do high levels of integration into the world economy (as measured by our index) decrease a state's ability to manage its public finances?' So here we have it: a research question which will enable us to investigate a narrower, more specific terrain relating to the broader question about globalization and the basis of the nation state.

There are two other issues involved in making a question researchable. The first has to do with whether you are asking the *kind of question* you want to ask. The second issue has to do with the inclusion of *errors of reasoning* in the statement of a question. We discuss both of these issues next.

Types of questions: getting clear about what you want to know

Once you think you have a question, and before proceeding any further with the research process, it is worth giving some thought to what your question is asking and what you will need to do in order to answer it. Is your question, as stated, asking for something different from what it is that you really want to know and do?

In order to get clear about what you are asking and want to know, it is useful to consider various types of questions that researchers ask. Not all types of research question will be appropriate to the requirements of your dissertation or other type of project as defined by your department. For instance, some departments may require that a dissertation be based on an explanatory question. However, though the requirements of your department may limit your choice among different types of research questions, it is important to be aware of, and understand something about, all types of questions. This will help you to formulate

BOX 4.6 Questions, What You Want to Know, and What You Will Do

Type of question	What you want to know	What you will do
Descriptive	The characteristics of something or of how it works or behaves	Describe the characteristics of something; model how it works or behaves
Explanatory	The causes of something that has occurred or is happening	Explain what factors or conditions are causally connected to a known outcome
Predictive	The future outcome of current conditions or trends	Predict what outcome will occur as a result of a set of known factors or conditions
Prescriptive	The things that can be done to bring about some outcome	Prescribe what should be done to prevent something from happening or to bring something about
Normative	What is best, just, right, or preferable and what, therefore, ought to be done (or not done) to bring it about (or prevent it)	Adjudicate among different understandings of how something should be, or what should or ought to be done, by considering the arguments of others, and submitting rational reasons for one's own

your question and be clear about what you will need to do to answer it; and it will also aid you enormously in making sense of the confusing, often conflicting, views that you find in the literature on your question. Consider the types of questions listed in Box 4.6. Each type of question is asking for a particular type of information. For instance, a descriptive question is one that asks for information about the characteristics or behaviour of something. This is perhaps stating the obvious. But by making this explicit you will be able to consider whether the question you have formulated is asking about the thing that you are really interested in knowing. Also, by making explicit what kind of question you are asking and what that question is asking for, you can begin to think about what you will be required to do in order to answer it. Let's consider each of the types of questions shown in Box 4.6.

Descriptive questions

Descriptive questions are concerned with the characteristics of what has happened, what is going on, or what exists, and with the characteristics of how something behaves. Descriptive questions often ask 'Who, What, Where, and When'. These can be research questions if they are significant and researchable, and if it can be shown either that people disagree about their answer or are, in some way, wrong about the answer on which most of them agree. For example, *who* killed Pakistan's Prime Minister, Benazir Bhutto? *What* national security doctrine guided US foreign policy under George W. Bush? *When* did the 'take-off' to industrial development occur in England? *Where* did nationalism, or the nation state model, first emerge? But descriptive questions come in a variety of forms. The question 'Have the foreign policies of EU member states become Europeanized?' asks whether we can *describe* the foreign policies of EU member states as having become 'Europeanized'. 'Has class diminished in importance as a basis for voting?' is a question which asks whether,

in *describing* the bases of voting preference, class is less important than it previously was. The question, 'What has been the impact of structural adjustment programmes on third world countries?' is asking for a *description* of the impact of these programmes. When we ask 'What proportion of people hold various opinions?', we are asking a *descriptive* question for which public opinion polls provide an answer. Some descriptive questions can be answered rather easily, as they have a clear factual answer that is readily accessible and uncontested. If so, they wil not require sufficient research to make them 'research questions'. Others, however, are not easily resolved or may be the subject of much contestation and so can provide much scope for theoretical and empirical investigation.

Explanatory questions

Explanatory questions generally ask about *what is causing or has caused* an outcome, or *why* something exists or has happened: 'What accounts for the resurgence of "Hindu nationalism" in India in recent years?' 'Why are British citizens generally less favourably disposed towards the EU than their counterparts in other member countries?' 'Why was there a reversal of democratization in Georgia?' These questions are answered through identifying the outcome to be explained (dependent variable) and the factor or factors thought to be connected with it (independent variables), and then showing what the nature of the connection or relationship between them is.

Explanatory and descriptive questions are both concerned with connections or relations between two or more variables. But while a *descriptive* question might ask what relationship exists between gender and voting preference, or of educational achievement and political participation, an *explanatory* question will be concerned with why that relationship exists. So, for instance, a descriptive question might ask what proportion of males and females say they would vote for a Democratic or a Republican candidate in the next US presidential election, while an explanatory question might ask why more women say they will vote for a Democratic candidate in the next US presidential election.

But there is not always a firm distinction between questions which ask for descriptions and those which ask for explanations. A question that asks about the causes of 'ethnic conflicts' is already suggesting something about their causes ('ethnicity'); and an elucidation of these causes will, in turn, confirm our description of them as 'ethnic' conflicts. The answer to the question of whether people with a university education are more apt to vote than those with only a secondary school education, offers both a description of those who vote *and* a possible explanation (level of eduction) for voting (White and Clark 1983: 23).

It appears to many people that explanatory and descriptive questions are more valued by, respectively, positivist and interpretivist researchers. This is because positivists tend to stress the pursuit of law-like explanations as the goal of social science research while interpretivists tend to place emphasis on interpretation (see Chapter 2). But both are concerned with description as well as with explanation. Where they differ is in their conception of what constitutes explanation. The difference, as Iain Hampsher-Monk and Andrew Hindmoor point out, is that positivist explanations are 'external' in the sense that they tend to work via correlations or deductions on the basis of *ascribed* reasons, and so need not concern themselves with actors' understandings of the world; while interpretive explanations are 'internal' in that they are concerned with the world of meanings inhabited by the actor (Hampsher-Monk and Hindmoor 2009: 48).

Predictive questions

Predictive questions are concerned with the likely effect or outcome of something or the trajectory of existing trends. In our field, researchers tend to be concerned with questions about the outcome of government policies: 'What will be the effect of further intervention on behalf of opposition groups, or to bring about regime change, in Syria?' 'What impact will the global energy transition (from oil to new forms of electricity) have on political regimes in the Middle East?' 'What will be the consequence of nationalizing insolvent firms?' Researchers are also concerned about the outcome of trends: 'What impact will increasing urbanization have on Egypt's unemployment rate?' 'Will China and Japan develop more cooperative relations in the next twenty years?' 'Will ethnic conflict increase in European cities in the next twenty years?'

Predictive studies use explanations to speculate about what is likely to occur or be the case in the future. So, for instance, if we analyse evidence and find that event A causes or explains event B (an *explanation*), we then can *predict* that if A continues to increase, we are likely to have a greater amount of B in the future (White and Clark 1983: 23). Predictive questions require you to develop an argument that if certain conditions or circumstances prevail, a certain outcome is likely to occur or come into being.

Prescriptive questions

These are questions that ask about what we should do, or about what is right or wrong, or good or bad. A prescriptive question asks: 'How can we bring about X (or prevent Y and Z)?' Researchers in our field are concerned with prescribing what governments should undertake to do in order either to bring about or to prevent something. 'What should the US do to bring about an end to the war in Afganistan?' 'What steps should the international community take to reduce human trafficking?' 'What can the British government do to stabilize its housing sector?' While prescriptive and normative questions are often closely connected, they are not necessarily the same. Prescriptive questions are more concerned with ascertaining facts needed to solve political problems. The question, 'Is democracy the best form of government?' is a normative question; but questions about what governments can do to expand political participation, or to ensure sound electoral processes, or to facilitate peaceful transfers of power are prescriptive questions.

So, questions that address what governments can do to bring about or prevent some specific outcome or policy goal are prescriptive questions. In undertaking research to answer these questions, a number of policy components will need to be analysed. These are shown in Box 4.7.

Your argument might involve agreement with the policy goal(s) pursued, but disagreement with the policy instruments being used. You might agree that the chosen policy instruments being used are the only or best ones available, but argue that they could be used differently. Prescriptive arguments can be investigated and defended in a variety of ways. For instance, researchers can use historical case studies of situations similar to the situation for which a policy is being prescribed in order to demonstrate that the recommended prescription is in some way better than the possible alternatives.

BOX 4.7 Policy Analysis

1. Identify the goal(s) the government is attempting to achieve.
2. Identify the policy instruments available. The identification of policy instruments can be made, first, without the burden of constraint analysis (see (3))—that is, in terms of what would be a logical means of achieving the goal.

   Resources or instruments for achieving policy goals might include the following:

   A. *Diplomatic*: threats and promises of various sorts, bargaining.
   B. *Military or paramilitary*: organized force and violence.
   C. *Political*: overt and covert activities of political support, political subversion.
   D. *Social*: technical aid, community development, training programmes.
   E. *Economic*: gifts and loans of funds to economic actors; channelling of investments, quotas, and restrictions. (Note: the withholding of economic aid is also a policy instrument.)
   F. *Informational*: propaganda directed at elites and groups.
   G. *Cultural*: exchanges of intellectuals, scientists, artists, students, etc.
3. Identify the various constraints that may operate on the resources available to policy-makers. These constraints narrow the range of possible alternatives open to the policy-maker. The constraints to pursuing any given policy option may include one or all of the following:

   A. The international environment
   B. The internal balance of forces in the country:
      -political instability;
      -domestic constituencies;
      -demographic factors.
   C. Constraints on political leaders which limit their ability to take action, on their organizational capacity, on the resources under their control, etc.
   D. Governmental agencies.
4. Evaluate the costs and benefits to the country of the available policy options for pursuing its goals. The costs incurred in undertaking or sustaining a line of policy include the money, people, and material employed or lost; the time, energy, or attention expended. There may also be unintended negative side effects of pursuing a chosen policy. A policy which might achieve its immediate goal may at the same time lay the foundation for future problems. In addition, there may have been important 'opportunity costs' incurred by the policy—that is, other things that might have been done or sought, but that were subsequently ruled out because they were in contradiction with the line of policy, or because the resources needed to seek them were pre-empted by the chosen policy.

Normative questions

While prescriptive questions are concerned with identifying the best means to given ends, normative questions are concerned with determining what the ends themselves should or ought to be. A normative question asks: 'What is best, just, right or preferable and what, therefore, ought to be done (or not done) to bring it about (or prevent it)?' 'When, if ever, is war and political violence justified?' 'What makes a government or state legitimate?' 'Are there transcendent or universal human rights?' 'Should human rights limit the way states can treat their citizens?' 'On what basis should trade-offs between liberty and security be decided?' 'What constitutes political equality, and does it require economic equality?' 'To what extent should liberal democracies tolerate illiberal cultures?'

As we pointed out in Chapter 1, political researchers have tended to define a sharp distinction between 'empirical' and 'normative' research. However, as we also pointed out, this distinction appears to have decreasing relevance for many scholars. Chapter 5 will address this issue in some detail.

Box 4.8 shows different forms that descriptive, explanatory, predictive, prescriptive, and normative questions take.

Unanswerable questions: fallacies in framing research questions

We have established that a question must be sufficiently narrow or specific to permit empirical investigation, and that you should be sure that, in the way that you have formulated it, it is a type of question that asks what you are really interested in answering.

BOX 4.8 Types of Questions

Your question	What you will do
What is terrorism?	**Describe** the characteristics of terrorism.
How are terrorists and their actions similar to or different from traditional political or military actors and actions?	**Describe** the characteristics of terrorism and those of other forms of violence, and compare them.
What are the consequences of terrorism?	**Describe** what the consequences of terrorism have been.
How might we expect various audiences and people to react to terrorism?	**Describe** how different types of individuals/groups react to terrorism.
What type of individual becomes a terrorist?	**Describe** the types of individuals who become terrorists.
Why do people or groups engage in terrorism?	**Explain** why people engage in terrorism.
What factors act as a stimulus or facilitator to terrorism?	**Explain** the factors that stimulate or facilitate terrorism.
How do the media affect our view of terrorism?	**Explain** how media content affects views of terrorism.
Where are future terrorist attacks likely to occur?	**Predict** that terrorist attacks will occur as a result of a particular set of factors or conditions.
Will terrorist attacks increase over the next decade?	**Predict** the future trajectory of terrorist attacks as a result of a particular set of factors or conditions.
How can the ideas or actions of terrorists be influenced?	**Prescribe** what can be done to influence the ideas or actions of terrorists.
What is the most effective way that governments can respond to terrorism?	**Prescribe** policies that will enable governments to respond most effectively to terrorism.
How can governments control terrorism while maintaining democratic freedoms?	**Prescribe** policies that will enable governments both to control terrorism and to preserve democratic freedoms.
When, if ever, are war and political violence justified?	**Make explicit** the moral implications of different points of view concerning when, if ever, war and political violence are justified; and **present reasons** why your own point of view is to be preferred.

However, a question that satisfies these conditions may still come up short if it contains errors of logic, or 'fallacies' that would generate false answers or conclusions. The most common errors of this sort are when a question is formulated in a way that:

1. 'begs' another question;
2. presents a false dichotomy;
3. asks about a fictional event;
4. is metaphysical;
5. is a tautology.

An example of each of these fallacies should suffice to make clear the sort of error it entails and the kind of question that results.

Consider this question: 'Why was American slavery the most awful that was ever known?' As the historian David Hackett Fischer points out, this is an explanatory question (a 'why' question) that begs a second (descriptive) question: the question of whether American slavery was 'the most awful that was ever known' (1970: 8). This second question '*goes begging*' in the sense that it is not explicitly asked and answered, but is simply assumed. In other words, the question is formulated in a way that really poses two questions, but is explicitly asking—and intending to answer—only one of them.

A question that is constructed so as to present a *false dichotomy* is one that forces us to choose between two answers that are neither mutually exclusive nor collectively exhaustive. Fischer describes what happens when questions such as these appear on an undergraduate essay examination: confronted with a question of this sort, the 'disgusted undergraduate' is forced to make a choice between 'unappetizing alternatives, or perhaps to combine them in some ingenious paradoxical contrivance of his own invention' (1970: 10). Some of Fischer's examples will probably remind you of countless others: 'Napoleon: Enlightened Statesman or Proto-Fascist?'; 'Jacksonian Democracy: Myth or Reality?'; 'Plato: Totalitarian or Democrat?'

Then there is the fallacy of '*fictional questions*'. Consider this example: 'Would US President Franklin Delano Roosevelt have decided to drop atomic bombs on Japan had he still been in office in August 1945?' The problem with this question is that there are no secure grounds on which to base an answer to it. We might be able to contrast the views of Roosevelt with those of his vice president, Harry Truman, who succeeded him and whose decision it was to use these weapons. But how can we know that, if he had still been president, Roosevelt would not have done precisely what Harry Truman did? Empirical research cannot answer this question.

We encounter the same problems with the fallacy of *metaphysical questions.* These are questions that attempt to 'resolve a non-empirical question by empirical means' (Fischer 1970: 12). Consider this question: 'Was the First World War inevitable?' How can you answer a question about inevitability through empirical research? Using words like 'inevitable', 'unavoidable', and 'inescapable' commits you to an argument that goes beyond what your research can establish.

Finally, there is the problem of *tautological questions*. These come in a variety of forms. The most common consists of the statement that 'things that are X are in fact X'; or 'X is X'.

Consider this question: 'Was George W. Bush unsuccessful because he was moving against the tide of history?' We would not know whether he was 'moving against the tide of history' except for the fact that he was unsuccessful. So this question asks if he was unsuccessful because he was unsuccessful. Here's another example: 'Did the assertion of separate national movements cause the centrifugal tendencies (i.e. the dispersal of political power away from the central government) that led to the dissolution of Yugoslavia?' This asks if centrifugal tendencies were the cause of centrifugal tendencies. If you ask, 'Do mainstream Americans believe that racism is unacceptable?', and further discussion reveals that 'finding racism unacceptable' is what you mean by 'mainstream', then what you are asking is whether mainstream Americans are mainstream.

'Researchability': a final consideration

To be researchable, your question must be within your ability to answer it. You need to consider the main practical issues involved in answering your question, including the availability and accessibility of data, and additional resources and training you might need. Will you be able to **access** people, statistics, or documents from which to collect the data you need to address the question fully? Can this data be accessed within the limited *time* and *resources* you have available to you? How long would you need to complete the research necessary to answer the question? What ethical constraints might be at issue? The answers to these questions (which we will consider further in Chapter 5) may lead you to further refine, or redefine, your research question.

So far, we have considered (1) the need to formulate a question so that it relates to an issue or area of inquiry that is significant for the field of politics; and (2) the various considerations that ensure that a question is researchable. These considerations relate to two of the three requirements for a research question. In the next section, we address the third requirement of a research question: that the question does not yet have a definitive answer. The next section explains how the 'literature review' enables you to establish this.

The literature review

The commonly used term, 'literature review', sounds like something mundane and mechanical: an obligatory chore contrived to ensure that you have done some reading on your topic. This is unfortunate; and it is also misleading. A review of the existing literature is not only a crucial part of the research process, it is where you demonstrate your ability to engage *analytically* with politics, where we hear *your* voice, and find out how you think and reason.

A literature review has a dual function. First, by explaining why, and to what extent, a definitive answer to your question does not yet exist, it demonstrates that your question meets the third requirement for a research question. Second, and in the process of performing the first function, it prepares the ground for the elaboration of your own argument. Box 4.9 shows how asking and answering a few key questions about the literature advances each of these functions.

BOX 4.9 The Dual Function of a Literature Review

1. Establish that a question has not yet been definitively answered, or that no one else has done what you are proposing to do, by asking:
   A. Who else has asked and answered your question? What (range of) answers have been given to this question in the existing literature?
   B. What is missing, irrelevant, defective, unwarranted, or ambiguous in the existing literature on your topic?
2. Set the stage for your own argument, by asking:
   C. What are the positive elements in the current literature? What in the literature can you highlight, underline, expand, extend, improve, build upon, or continue?
   D. What needs to be done? What requires further investigation or re-thinking? What would provide a better understanding of the matter?

The literature review demonstrates that the subject has not been adequately dealt with in the existing literature; and it explains how you intend to address this state of affairs, either by doing something that has not already been done that promises to yield further insight into the question, or by doing what *has* already been done, but better. This may require you to show that, though the question might already be the subject of mountains of articles and books, the answers that have been offered to it are in some way incomplete or flawed. What you will need to do is to identify other researchers who have asked and answered your question, and say what about their answers is wrong, incomplete, or weak; and what about them is valuable. It is on the basis of both their positive and less positive elements that you build your case for what you yourself propose to do to contribute to the literature.

This is the second function of a literature review: to set the stage for a statement of how your research is going to address the weaknesses of existing studies, and/or underline, expand, extend, improve, or build upon their strengths. Your review of the literature that offers answers to your research question, and your assessment of its weaknesses and strengths, paves the way for the presentation of your own argument. Before discussing how to develop a literature review and its association with your argument, we should note a concern that students sometimes raise. Students sometimes think that there is no literature relating to their topic, and so nothing they can use as a basis for a literature review. The answer is that there *is* literature. There may not be books and articles that address *exactly* the same topic, but there will always be related research that will provide a conceptual and theoretical context for your research. Think about the broader, top-of-the-vase, questions which frame, and are related to, your research question; ask yourself: 'What is my question a case of? Is there a broader literature on this type of phenomenon?'

Let's consider the way you develop a literature review and how it relates to the development and presentation of your own argument. There are three stages involved in producing a literature review. In the first stage, you read, follow leads to other promising books and articles, and then read some more. The key here is to read *analytically*—that is, to ask a lot of questions about what you are reading, and to also focus on getting a sense of the different debates, theories, perspectives, and arguments relating to your question. The goal is to eventually bring some kind of organization to the literature, by summarizing how the literature

breaks down into different theoretical perspectives, entry points, or arguments. This is the second stage of developing a literature review. Finally, and as a result of the previous two stages, your own argument will begin to come into focus. You should now have answers to the questions listed under 'The literature' in Box 4.1 ('Outline of Research Components'). At this point you are ready to write the literature review: a discussion of the best research and writing on your question organized in a way that sets the stage, and helps to make the case, for your own argument.

Let's look more closely at each of the three stages in the development of a literature review.

Stage 1: reading the literature

We have said that you should approach the literature *analytically*. This means that, as you evaluate the major arguments or explanations in the existing literature on your question, you should be asking a lot of questions. You should query the questions, answers, methods, and conclusions of each contribution to the literature you read; and scrutinize their ideas, definitions, concepts, information, or logic. Ask yourself what is missing, irrelevant, defective, unwarranted, or ambiguous. There are numerous questions that can be asked about each aspect of the material you read. Consider, for instance, the questions, below:

1. What is *the author's central question* or subject of inquiry? How does the way the question is asked make it different from those which others have asked? How does it illuminate, or obscure, crucial dimensions of the phenomenon investigated? What is the author's *purpose* (to describe, explain, predict, prescribe)?

2. What is *the author's answer or argument*? What factor or factors is the author's analysis highlighting? Does the author define these factors clearly? (What is meant by 'technology', 'rapid change', 'instability'?)

a. How does the author's argument or explanation differ from others? (Note that the title of the book or article is usually designed to communicate what differentiates this particular treatment of the subject matter from other books and articles; and the chapter titles and section headings provide clues to how the overall argument is developed.)

b. If the purpose of the inquiry is to explain something, what are the author's major hypotheses? Are they developed *deductively* from a more basic theory? Are they developed *inductively*? What is the strength, nature, and generality of the relationship between the independent and dependent variables?

c. What assumptions underlie the account of events that the author presents? Are the assumptions sound? Does the argument assume things that need to be demonstrated?

d. Does the basic argument make sense? Is the chain of reasoning sound? If the author moves from one claim to another that supposedly follows from the first, does the author demonstrate a clear relationship between the two claims? Are there any leaps of logic, inconsistencies, circular arguments, or contradictions within and among the parts of the argument or the subsidiary arguments?

3. What *methods* are used to demonstrate the argument? Are there any shortcomings or flaws with respect to the methods used? Is the time frame too short, or artificial; or is the spatial domain too limited or artificial to generate sound conclusions? Are key definitions ambiguous, or too narrow or broad?

4. What *evidence* is presented in support of the claims that are made? Is the evidence well-selected, or is it arbitrary? Does the author use many different kinds of examples, or ones that are chosen simply because they support a particular contention or case? Does the author address examples of what might be considered a 'hard case' for the argument—counter-examples that might undermine the central claims? Can you think of an anomaly—a case that doesn't fit with or isn't explained by the theory? Has the author ignored, suppressed, overlooked, or slighted relevant evidence, alternative data sources, or other interpretations of the data?

5. Do the *conclusions* fully draw the implications of the overall argument, all parts of the argument, and subsidiary arguments?

Stage 2: summarizing the literature

In Stage 1, you read widely, but selectively, in your topic area and asked questions of everything in the material you read; and you will also have taken *careful notes* all along the way! It is time, now, for you to bring some organization to this literature—to summarize it in a way that identifies the major themes, issues, arguments, positions, or perspectives that constitutes it and that highlights their major points of contact and division.

You will not need to summarize all the literature relating to your topic, but only the literature that directly addresses the specific question you are asking. Depending on your question, it might be appropriate to summarize the intellectual progression of the field of study that relates to your question, including major debates. Alternatively, your question might require you to focus on sources that address the relevant subject matter within a certain time period. But, whatever you chose to include in your summary of the literature, it should include the main conclusions of the most important and influential arguments and ideas in the existing literature relating to your question. The best way to find these conclusions, and to make decisions about how to 'bring organisation' to the literature, is by asking and answering the following questions:

- What are the principal positions that experts take with respect to causes, consequences, or solutions? Do existing studies tend to divide between two or three different approaches or perspectives?
- Are there different schools of thought? If so, who are the major disputants, and what points of difference lie between them?
- What are the aspects of the topic about which experts appear to agree? What are the principal insights that have emerged from the literature?

Answering these questions will enable you to organize and summarize the literature. It will also raise a number of important additional questions:

- What puzzles or problems has the literature left unsolved?
- Is there an aspect of the field that is missing?

- Do the arguments and theories in the literature tend to focus on examples or cases drawn from one part of the world?
- Have they been/can they be applied to other parts of the world?

These are the sorts of questions that will move you towards the third and final stage of the development of a literature review.

Stage 3: what still needs to be done? Setting the stage for your own argument

You have read, and you have thought creatively and critically about the arguments in the literature and the findings of previous research relevant to your topic (stage 1). You have been able to bring some organization to the literature by breaking it down into the major points of contact and division among the various perspectives and arguments that it contains (stage 2). In the process of doing this, you have identified strong and weak elements in the literature, and this has led naturally to a consideration of what might be done to build on those strengths and overcome those weaknesses. Consequently, in the process of analysing the positive and less positive elements in the arguments others have made, your own argument has begun to come into focus. You are now ready to move to the final stage in the development of a literature review: writing a discussion of the existing literature that presents *a recapitulation of your own process of evaluation and critique* of it and *the process by which your own argument began to take shape*. In sum: your 'literature review' discusses the best literature on your question in a way that shows how a reasoned analysis of its strengths and weaknesses leads naturally to a consideration of what you propose to do as a contribution to resolving the question. It leads, naturally, in other words, to the introduction of your own argument.

Here are some basic considerations and ground rules to guide the writing of your literature review:

1. A literature review should itself *develop an 'argument' about the literature*. It should begin with a statement of that argument and use evidence to back it up. The argument of your literature review is just like any other argument: your interpretation of the material you discuss must be backed up with evidence to show that what you are saying is valid. The literature review should contain an introduction, the body of the review containing the discussion of specific sources, and a conclusion.

2. A literature review is *not* an annotated bibliography—a list of sources with a discussion of each one of them, one at a time. Instead, you should *organise your discussion, not around the sources themselves, but around ideas, themes, theories, or issues* that enable you to advance the argument of your literature review.

3. The number of sources you discuss depends on how many you need in order to persuasively demonstrate the argument of your literature review. You must satisfy a critical reader—for instance one of the authors whose work you are challenging—that your argument is based on considerable thought and study and that you know what you are talking about. Comprehensive knowledge of the literature on a topic is essential to research on that topic; and the depth and breadth of knowledge that your literature review exhibits is what will give you credibility as a researcher.

4. Address the best arguments, including those with which you agree and those with which you disagree. Don't leave out good arguments in order to make a case that there are no good arguments along the lines of what you yourself may want to develop. Don't downplay or misrepresent the sophistication and implications of the arguments advanced by a strong opponent; and don't go after a weak opponent (a 'strawman'). It may be the case that the only opponents *are* weak. Be sure, however, that you haven't overlooked a stronger one.

5. The types of sources that should be included (books, journal articles, websites) will depend on your topic. You are concerned with the state of the best current literature and previous research on this question. Focus on the articles and books that are most influential and widely cited. Concentrate your efforts on relevant articles published in the most credible research journals dealing with the general area of inquiry related to your question. Ideally, these should be research journals that use a blind-review system. In a blind review, authors submit potential articles to a journal editor who solicits several reviewers who agree to give a critical review of the paper. The paper is sent to these reviewers with no identification of the author so that there will be no personal bias (either for or against the author). Based on the reviewers' recommendations, the editor can accept the article, reject it, or recommend that the author revise and resubmit it. Articles in journals with blind-review processes can be expected to have a fairly high level of credibility. With regard to the use of websites, please consider the issues in in Box 4.10 relating to the use of internet sources.

6. Select only the most important points in each source that relate directly to the argument of your literature review.

7. A good literature review is one in which the material is thoroughly 'processed': organized according to an overall argument and around key issues or themes, and rendered in the writer's own voice. While the literature review presents other people's ideas, your voice and your own perspective should predominate. Use quotes sparingly and maintain your own voice by starting and ending the discussion of different issues, themes, or individual sources with your own ideas and words, using the sources to support what you are saying. Find models: read the literature reviews in books and articles written about your area of interest to get a sense of what style and organization is effective.

BOX 4.10 The Use of Internet Sources

The internet is a fast and easy way to access information. It is a source of good information, but it is also a source of half-truths and complete falsehoods. Consequently, you will have to figure out what is a legitimate source and what is not. In general, authoritative sources are ones that are known to verify their facts and information to a reasonable extent before publication. Reputable academic journals, for example, normally only publish articles that have been at least twice blind-reviewed by experts in the field who agree that the information is essentially correct and that the material is usefully contributing to a body of knowledge. Major news sources, which are liable under libel legislation and have an interest in being trusted by their readership, also attempt to meet a high standard of journalism.

There are, however, millions of publications and websites which are geared towards presenting opinions and rumours rather than research and certifiable facts. These include blogs, 'quasi' news, and gossip sites. Essentially, the key comes down to verifiable information. Can the source verify the facts given—or can the authors prove their case? In the case of blogs and opinion-based websites or user content-generated sites, probably not. Examples of academic or authoritative sources online include articles from major sources (BBC, CNN, AFP), government websites, and reputable non-governmental organizations (NGOs) (e.g. International Committee of the Red Cross, Amnesty International, National Union of Journalists). *Information sources such as Wikipedia, About.com, or Infoplease.com are not academic sources.* Wikipedia and other sites like it do not provide sufficient monitoring of what information goes up, and frequently the information is wrong or misleading.

We have discussed what a literature review is (and is not), what it consists of, and how to go about developing one. Use this discussion to guide you through the process of developing a literature review. But also take a look at the examples of literature reviews included in the Online Resource Centre. Once you've looked at these examples, you will be able to identify literature reviews in the articles and books you read, and consider which you find most readable and helpful, and why.

Conclusions

This chapter has been devoted to the first step in the research process: finding and formulating a research question. We have discussed the following:

- Three requirements of a research question. We have said that for a question to be a *research question* you will need to:

  1. articulate a rationale for pursuing research on the question in terms of its significance for a broad issue or area of interest within our field;
  2. formulate the question so that it is researchable—so that it can be answered through research; and
  3. show that the question has not already been definitively answered.

- How to go about meeting these requirements. We introduced a visual tool—the 'research vase'—which can help you to conceptualize the first two requirements of a research question. To meet the third requirement of a research question you must show how others who have asked this question failed in some way to provide a definitive answer to it. The literature review performs this function. It identifies who else has asked and answered your question, and what (range of) answers are found in the existing literature relating to it. It highlights the positive elements in the current literature, and makes an argument about what needs to be done in order to provide a better answer to the question than currently exists.
- Different types of questions and what each type commits you to doing.
- The logical fallacies that sometimes find their way into the statement of a question and that lead to false conclusions.

Once you have formulated a research question, you have completed step one of the research process. You will then be ready to move on to step two: developing an answer to your question. This is the focus of Chapter 5.

Questions

1. What makes a question 'researchable'?
2. How do you formulate a question so as to maximize its generality?
3. How might errors of logic in the formulation of a question generate false answers or conclusions?
4. What is the difference between a survey or summary of the literature and a 'literature review'?
5. How are the requirements of a research question related to the rationale or justification for pursuing research on a question?

Guide to Further Reading

Browne, M. N. and S. M. Keeley (1994), *Asking the Right Questions: A Guide to Critical Thinking*, 4th edition (New York: Prentice Hall).
The authors provide a wide range of questions that can be used to interrogate the existing literature on a topic.

Buttolph Johnson, J., H. T. Reynolds, and J. D. Mycoff (2008), *Political Science Research Methods* (Washington, DC: CQ Press), chapter 3.
This chapter discusses the steps involved in specifying a research question and developing a hypothesis.

Geddes, B. (2003), 'Big Questions, Little Answers, How the Questions You Choose Affect the Answers You Get', in *Paradigms and Sand Castles* (Ann Arbor, MI: Michigan University Press), chapter 2.
This chapter makes the case for caution in using the literature in our field as a source of research questions, and argues that 'looking for gaps' in the literature may not be a good strategy for finding research questions.

Hedrick, T. E., L. Bickman, and D. J. Rog (1993), 'Defining the Focus of the Research', in *Applied Research Design*: A Practical Guide (London: Sage Publications), chapter 2.
This chapter provides practical advice on getting clear about the problem or issue that interests you and identifying the research questions that can address it.

Hoover, K. and T. Donovan (2004), *The Elements of Social Scientific Thinking*, 8th edition (Belmont, CA: Thomson-Wadsworth), chapter 3.
The authors present strategies for narrowing the focus of a research question, and discuss hypotheses, operationalizing concepts, and how to relate research questions to broader areas of concern.

Kahane, H. and N. Cavender (2005), *Logic and Contemporary Rhetoric: The Use of Reason in Everyday Life*, 10th edition (New York: Wadsworth Publishing), chapters 1 and 2.
These chapters discuss in detail common errors of reasoning or fallacies that prevent clear thinking and the formulation of good research questions.

King, G., R. Keohane, and S. Verba (1994), *Designing Social Inquiry* (Princeton, NJ: Princeton University Press), chapter 1.
The authors discuss the value of formulating a research question so that it both addresses a real-world issue and contributes to scholarly literature in a given field. They suggest a variety of ways that the literature can be used as a source of developing a research question.

References

Dunleavy, P. (2003), *Authoring a PhD: How to Plan, Draft, Write and Finish a Doctoral Thesis or Dissertation* (Basingstoke: Palgrave Macmillan).

Fischer, D. H. (1970), *Historian's Fallacies: Towards a Logic of Historical Thought* (New York: Harper).

Geddes, B. (2003), *Paradigms and Sand Castles* (Ann Arbor, MI: Michigan University Press).

Hampsher-Monk, I. and A. Hindmoor (2009), 'Rational Choice and Interpretive Evidence: Caught between a Rock and a Hard Place?', *Political Studies* 58(1): 47–65.

King, G., R. Keohane, and S. Verba (1994), *Designing Social Inquiry* (Princeton, NJ: Princeton University Press).

Pollins, B. M. (2007), 'Beyond Logical Positivism: Reframing King, Keohane, and Verba', in R. N. Lebow and Mark Lichbach (eds), *Theory and Evidence in Comparative Politics and International Relations* (New York: Palgrave/Macmillan), 87–106.

Shively, W. P. (1989), *The Craft of Political Research*, 2nd edition (Englewood Cliffs, NJ: Prentice-Hall, Inc.).

White, L. G. and R. P. Clark (1983), *Political Analysis: Technique and Practice* (Monterey, CA: Brooks/Cole Publishing Co.).

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

5 Finding Answers: Theories and How to Apply Them

Chapter Summary

Chapter 4 was devoted to a discussion of how to find and formulate a research question. We turn now to a discussion, which will be the focus both of this chapter and of Chapter 6: of how to develop an *answer* to a research question. In this chapter, we address the following issues:

- What are the requirements and components of an answer to a research question?
- What is a 'theory' and what is its role in social science research?
- What is a 'theoretical framework'?
- What is a hypothesis?
- What are the components of a hypothesis?
- What types of hypotheses are there and how do they guide different kinds of research?
- Why do you need, and how do you formulate, conceptual and operational definitions of your key terms?
- How do you go about answering normative questions?

Introduction

We have discussed *Step 1* of the research process: the formulation of a well-crafted research question (Chapter 4). We turn, now, to *Step 2*: how to develop a hypothesis or argument that answers it. We discuss the basic components and requirements of a hypothesis or argument, and where to find and how to formulate one.

What does answering a research question involve? The various considerations and tasks involved can best be understood in relation to three basic requirements. Different types of questions demand different types of answers, so the first requirement of an answer to a research question is that it be appropriate to the type of question that is being asked (see discussion in Chapter 4 and Box 4.8). The second requirement of an answer to a research question is that it makes a contribution to collective knowledge in a particular field of study. In other words, your answer must matter. So, as you develop an answer to your research question, you must ask yourself: Why should we care about this answer or argument? How does it contribute to the development of knowledge about politics?

You contribute to the development of knowledge in our field by relating your question and answer to existing theory. The reason for this is that in order to achieve progress in a field of study, research must be cumulative; and theory facilitates cumulative research. It links one study with another, and helps to ensure that the work of researchers enters into dialogue with and builds on the work of others. As we discussed in Chapter 4, researchers develop both a question about a topic and an answer to it, through a critical analysis of existing theory and research relevant to that topic. All research questions and answers are informed by a set of expectations derived from previous research and theories. Recall that the function of your 'literature review' is to explain the contribution your research will make. *Different types of research provide different contributions to knowledge.* What shortcoming in the existing literature on the topic will your research address? What contribution will it make to our knowledge of something? How, specifically, does it contribute empirically, conceptually, or theoretically to our knowledge of that topic?

In order to contribute to knowledge in a field of study, an answer to a research question must meet a third requirement: it must be clearly and fully *specified*. It must be specific with regard to (1) the factors or elements you think must be considered in order to answer your question; and (2) how you think these factors or elements are related to each other. Together, these factors and their relations constitute a '*hypothesis*': a reasoned, clearly specified hunch or expectation with which you begin your research and which helps to guide and focus your research. An answer or argument in response to a research question consists of a hypothesis and an investigation of it. Developing a hypothesis encourages you to be very precise about how your answer relates to those that others have offered to your question.

The term 'hypothesis' is often treated as applicable only to quantitative research and to a specific prediction about the nature and direction of the relationship between two 'variables' (we will discuss the term 'variable' presently). But we use the term 'hypothesis' in a far more expansive way. What we mean by 'hypothesis' is 'a hunch, assumption, suspicion, assertion, or idea about a phenomenon, relationship, or situation', with which research begins and which becomes the basis of inquiry (Kumar 2005: 74). Hypotheses can either be tested with evidence (confirmatory research), or operate as a guide to a process of discovery (exploratory research). *Confirmatory research* begins with a hypothesis and uses observations to test it. We begin with a statement, on the basis of a theory, of what we would expect to find, and then see whether what we expect is fulfilled. *Exploratory research* begins with a question and perhaps a basic proposition, probes its plausibility against various types of data, and eventually generates a hypothesis as a conclusion rather than as a preliminary to conducting the research itself (Schmitter 2008). In this type of research, we might therefore develop an answer through (1) a preliminary hunch; (2) an investigation; and then (3) a more concrete hypothesis. Whichever type of research you pursue, and irrespective of whether you are addressing a descriptive, explanatory, predictive, prescriptive, or normative question, thinking in terms of formulating a hypothesis can help you to clarify your argument and the kinds of evidence that will provide a meaningful demonstration of your answer.

As these statements suggest, we see a hypothesis as a basic infrastructural element whose function, in all research, is to help to make explicit the argument that is being developed. For all types of research, we think it is useful to formulate a 'working hypothesis'—an operational hunch about what you expect to find. What argument is motivating your research? What findings might be expected to result from it? By articulating in advance the contours and logic of the investigation, a hypothesis serves as a guide for your research. As we shall see, *different types of research are guided by different types of hypotheses*.

To further underline our point that a hypothesis is a basic element in all research, recall that in Chapter 2 we argued that, irrespective of whether researchers see themselves as positivists or interpretivists, practical investigation of research questions often leads them to undertake methodological tasks and research practices which, as a number of scholars have pointed out, are founded in the hypothetico-deductive method. This method, as we discussed in Chapter 2, involves the formation of *hypotheses* and the *deduction* of consequences from them. Researchers either 'assess whether the information they have gathered fits with the interpretation they have posited', or they assess whether 'competing interpretations fit with the facts they have gathered'; but, in either case, as Brian Pollins points out, 'they are practicing the hypothetico-deductive method' (2007: 100).

We hope that we are persuading you of the value of beginning research by framing a hypothesis as a basis for your inquiry. But we know that the idea that we should form a hypothesis about what we expect our research to reveal *before* we begin our research may seem, in practice, to be unrealistic or problematic. How, you might ask, can we formulate a statement of what we expect to find before we actually begin our research? Don't we need to have done at least some of the research in order to come up with a 'hunch' about what it might reveal? And shouldn't we expect that our view on this will, in any case, likely change (perhaps several times) as our research proceeds?

The issue that this raises has to do with an important distinction: the distinction between the *process* of research and its *presentation*. The *process* of research involves a dialogue between ideas (theory) and evidence. Most researchers move back and forth between theory and evidence—between theorizing something that is the focus of their research and mucking about in the 'dust of detail' (learning more about the specific facts of the case or question or issue, or the observations of other scholars that we treat as facts). As W. Philips Shively observes,

> one of the best-kept secrets in our field is that good researchers usually do not 'frame hypotheses' in any formal sense before they start to work, though they may have some operational hunches about what they expect to find . . . They play with data, immerse themselves in what other people have written, argue with colleagues, and think. (Shively 1989: 25)

As this chapter explains, the research process can be *conceptualized* as a series of steps. These steps, and the order in which they occur, constitutes a logic of inquiry. This logic of inquiry, if kept clearly in mind, will help to guide us through the research process and inform the many choices it requires us to make. And organizing the presentation of our research around this series of steps enables us to make a clear and compelling case for the reliability of our findings. However, *in practice*, the process does not unfold in this sort of linear fashion. We re-think our views as a result of learning from the research process itself in ways that can feed back to our previous choices and lead us to revise them. In practice, then, the research process is, not linear, but often circuitous.

Despite the messiness of the research process, as actually practiced, it seems reasonable to agree with Shively's recommendation that starting with 'clear and obvious procedures' that are 'more methodical and easier to apply for the beginning researcher' is a good way to learn how to do research (Shively 1989: 25–6). As students gain more experience, they will doubtless become more creative and develop their own process of work.

Political inquiry encompasses a variety of types and processes and possible starting points. But whatever the *process*, we suggest that the *presentation* of research makes clear the logic behind your reasoning and the assumptions upon which you are relying.

All types of research should be arranged carefully, systematically, clearly, and logically. What we read should be a coherent narrative whose storyline moves through research question, relevant literature, hypotheses, procedures and methods, findings, conclusions, implications.

The sections that follow will elaborate on all these points, using the three requirements of an answer to a research question we've just discussed. We discuss the components of, and answer to, a research question; approaches to developing hypotheses and the elements that constitute them (variables, and the relationships that can be established among them). We also consider the ways you must specify your argument or answer: the need to use terms that mean what you want them to mean, and that can be understood by others in precisely the way you want them to be understood; the necessity of providing a conceptualization (or *conceptual definition*) of your basic terms or variables, as well as an even more specific definition which identifies empirical referents for them (an *operational definition*). Chapter 6 will continue our discussion of how to answer a research question by focusing on research design and on data and sources. Box 5.1, which reproduces the outline of research components we presented in Chapter 4, indicates the elements that we have already covered, and those that we will be covering in both this chapter and Chapter 6.

BOX 5.1 Outline of Research Components

1. **The question** [a]
   - A. What do you want to know? What is the *central question/problem/issue/puzzle*?
   - B. Why do you think it is worth doing/knowing? What is *the rationale for pursuing research on this question*?
2. **The literature** [a]
   - A. Who else has asked and answered your question? What (range of) answers to this question are found in *the existing literature relating to it*?
   - B. What are the *positive elements* in the current literature? What in the literature can you highlight, underline, expand, extend, improve, build upon, continue?
   - C. What needs to be done? Delineate the crucial aspects of the problem requiring investigation. *What needs to be done in order to provide a better answer to the question than currently exists*?
3. **Your answer** [b]
   - A. *Theoretical framework*. What are the *theoretical elements and guiding assumptions of the study*?
     1. What factors or variables of the problem must be investigated in order to answer your central question?
     2. What is/are *your hypothesis/es* (how are these factors linked)?
        - a. What is the source of your hypothesis/es? What in theory would lead us to expect the relationship(s) you assert?
        - b. How would you demonstrate the relationships stated by the hypothesis/es?
     3. What is *the spatial/temporal domain of the study*? What is the rationale for defining this domain for the study?
   - B. *Data and sources*
     1. What are *the data relevant to demonstrating the relationships you hypothesize*?
     2. What *sources* are there *for these data*?

[a] Discussed in Chapter 4.
[b] Discussed in the present chapter and in Chapter 6.

Answers to research questions: general requirements

An answer to a research question (1) provides the type of answer appropriate to the question you are asking; (2) contributes to the development of knowledge in the field of politics by relating itself to existing theory and research on a topic; and (3) clearly and fully specifies its key elements and how they are related to each other.

What type of answer does your question require?

A research question, as Chapter 4 emphasized, not only initiates the research process, but is crucial to every step along the way. The kind of question you ask determines the type of answer you provide, the research you pursue (confirmatory, exploratory), and the means by which you pursue it. *You must provide the type of answer that your question requires*. If you were asked in an exam to '*Explain* why Iran and China have pursued different population growth policies', an answer designed to '*describe* what the key differences are in Chinese and Iranian policies used to address population growth issues' would earn you a low mark. The point, of course, is that the type of answer must be appropriate to the type of question. That means that, since, as we maintain, an answer to a research question entails an investigation of a hypothesis, different types of questions will require different kinds of hypotheses. So, let's review the different types of questions researchers in our field ask and then consider what sort of hypothesis might be formulated to order to answer them.

In Chapter 4 we identified five different types of question that researchers in our field ask: descriptive, explanatory, predictive, prescriptive, and normative questions. (Note that, as we pointed out, your department may not permit you to use all these types of question as the basis of a research or dissertation project; some departments may require that your dissertation be based on an explanatory question; other departments may allow other, or all, types of questions.)

There is a tendency to treat hypotheses and hypothesis-testing research as applicable only to explanatory and predictive questions. *Explanatory questions* are concerned with what is causing or has caused an outcome, or why something exists or has happened. An *explanatory hypothesis* advances a guess as to the cause of, or reason for, something by identifying what factors or conditions are connected to a known outcome. They state, for instance, that X is caused, or made possible, by Y. *Predictive questions* are concerned with what will occur or be the likely outcome of something. A *predictive hypothesis* generalizes from an understanding of current or past events to predict the future outcome of current conditions or trends. Using evidence on current events, accumulated knowledge of general patterns, and relevant theories and theoretical research, it starts with conditions that are thought to be causal, and predicts the resulting phenomena. A predictive hypothesis claims that factor X will cause event or state Y, that outcome Y will occur as a result of a set of known factors or conditions X, or that in conditions ABC, event or state Y will tend to occur.

While explanatory and predictive questions are types of questions most usually associated with hypothesis-driven research, we maintain that all research questions and answers are informed by a set of expectations (hypotheses) derived from previous research and theories, and that, for all types of research—including descriptive, prescriptive, and normative research—it is useful to formulate a 'working hypothesis' or operational hunch about what you expect to find.

Let's consider the sort of hypothesis that might be formulated in response to a descriptive question. *Descriptive questions* are concerned with the characteristics of what has happened, is going on, or exists; or of how something behaves. The function of a hypothesis in descriptive research is to select, based on an assessment of what we think will prove most relevant, just those factors from the world of facts and figures, people and events, that we think are most useful for directing and focusing our research. We cannot simply describe everything relating to a phenomenon. We must always select and organize, out of the universe of possibilities, those aspects we think will prove important to explore and highlight. We can start with a theoretical construct or theoretically informed notion, hunch, or expectation that will serve as a *descriptive hypothesis*; perhaps an ideal type—a set of characteristics that we expect to find in some relationship to each other in a given case. A descriptive hypothesis might state, for instance, that X has A, B, C characteristics and/or behaves in D, E, F ways.

Description often involves an effort to set up definitions and classifications, or to make sense of alternative conceptualizations of a topic. The basis of this type of descriptive analysis might be a hypothesis that something can be *meaningfully described*, or usefully defined, seen, or interpreted, *as something else*; that a given concept is useful for describing or understanding something; that we shall understand this thing better if we see it in this way, or if we interpret it as being divided into these three types. We might call this sort of descriptive hypothesis an *interpretive* hypothesis. When we state, for instance, that all X belong to class Y; or that X can be interpreted as Y, we are hypothesizing that we can make good sense of X if we classify it, or interpret it as representing or 'meaning', Y. Does 'democracy' or 'political participation' in China *mean the same thing* as these terms do in the United States or Peru? Can we interpret (meaningfully describe) older democracies today as facing a 'crisis of political participation'? The 'test' of this sort of hypothesis is showing the usefulness of describing or interpreting X *as Y*.

Hypotheses can also be formulated as a basis for addressing prescriptive and normative questions. *Prescriptive questions* ask *what we should do* to bring about, or prevent, some outcome; what course of action we should follow to achieve a particular objective or goal. A prescriptive question might ask, for instance, 'How can the "brain drain" be reversed in South Africa?' A *prescriptive hypothesis* states that to achieve desired end X, we should do Y. To investigate this kind of hypothesis, the researcher will inventory available options, weigh the pros and cons of each for achieving the desired outcome and, based on this analysis, advance an argument for a policy, using existing theoretical and empirical research to verify the argument's factual and theoretical assumptions. *Normative questions* deal with questions of what is right and wrong, desirable or undesirable, just or unjust in society. A *normative hypothesis* is an argument or suggestive proposition about *what ought to be* that is advanced for further discussion, analysis, and investigation. A normative hypothesis might state that the best type of X is Y; that the basis for deciding or justifying X ought to be Y, the legitimacy of X ought to be based on Y, or that X is just when conditions A, B, C are present. Investigation of these hypotheses would entail elaborating an argument about why this should be the case and establishing the plausibility and coherence of the argument. The 'test' of a normative hypothesis is the rigour of logic and internal consistency of the argument.

In sum, it is useful to formulate a 'working hypothesis' for all types of research, but the hypotheses that are formulated must be appropriate to the type of question being asked. Box 5.2 shows the different kinds of hypotheses that different types of questions will require.

BOX 5.2 Types of Questions and Forms of Hypotheses

Type of Question	▶ Form of Hypothesis
Descriptive	X has A, B, C characteristics and/or behaves in D, E, F ways.
Explanatory	X exists as a result of A, B, and C.
Predictive	If circumstances X and Y prevail, Z is likely to occur or come into being.
Prescriptive	In order to bring about X, do (or don't do) A, B, C.
Normative	X is preferable to Y and Z, therefore A, B, C ought to be done (or not done) to bring about X (or prevent Y and Z).

Some examples of different types of research questions and the hypotheses that might be offered in answer to them are shown in Box 5.3. Note that a hypothesis is stated affirmatively, that is, not as a question.

BOX 5.3 Research Questions and Hypotheses

Type of Question	▶ Research Question	▶ Hypothesis
Descriptive	What is the effect of negative advertising on political participation?	As the number of hours of negative ads watched increases, the probability that an individual will vote decreases.
Explanatory	What are the root causes of the Chiapas conflicts?	Conflict resulted from the deterioration of conditions for the indigenous population due to the drop in coffee prices in 1989, the reform of article 27 of the Constitution of 1992 to facilitate the commercialization of land, and the North American Free Trade Agreement which came into effect in 1994.
Predictive	Will China and Japan develop more cooperative, or more conflictual, relations over the course of the next twenty years?	Economic and security interdependence will increase over the next two decades, leading to an easing of historical tensions and greater cooperation between the two countries.
Prescriptive	What steps should the international community take to reduce human trafficking?	Within complicit or non-cooperative countries, tax treaties should be used to create internal pressure on the governments and on their wealthiest citizens who have the political power to force change.
Normative	Are there transcendent or universal human rights?	All rights and values are defined and limited by cultural perceptions. There is no universal culture, therefore there are no universal human rights.

Answers that contribute to the development of theory

We have said that an answer to a research question should contribute to the development of knowledge in the field of Politics and International Relations by relating itself to existing theory and research on a topic. So: what is a '**theory**', and why do we need it (what does it do for us)?

The world is complex, and a theory is an attempt to make sense of complexity. It does this by indicating that some factors are more important than others for explaining a given outcome, and showing that these factors were not randomly chosen but are related to each other. A theory is always a skeletal version of reality: it encompasses only a portion of the real world, deals with only a few variables, and makes simplifying assumptions to keep variables and cases manageable. A good theory is parsimonious: it attempts to explain as much as possible with as little as possible (parsimony). It is also generalizable: it can be applied to a variety of contexts or settings.

The nature and function of theory in social research can be brought more clearly into focus by contrasting theory-driven research with an ideal-typical 'non-theoretical' research enterprise. Consider, for example, a study of the causes of the First World War. The 'non-theoretical' researcher, unconcerned with either parsimony or generalizability, might choose to focus on any and all factors historians have recorded as present at the time of the outbreak of the First World War. Here, for example, are a dozen of the possibly hundreds of factors relevant to explaining the outbreak of the war (note how some of these factors are themselves 'theoretical', that is, the product of prior theorizing): (1) the international balance of power and the change in the configuration of the international system (a rigidified balance-of-power system, which heightened competition and tension, led to an arms race and a fear of losing allies); (2) British hegemonic decline and Anglo-German rivalry; (3) German nationalism and other attributes of the German state, which gave it a propensity for aggression; (4) Austria's struggle to survive and its effort to eradicate the threat of Serbian nationalism within its empire, and in the Balkans generally; (5) Russian geopolitical imperatives (its quest to gain control of the Dardanelles for access to the Mediterranean and thus to the world's commerce); (6) British fear of German control over the channel ports; (7) France's ambition to reacquire Alsace-Lorraine; (8) Germany's internal class relations, which caused the state to pursue foreign economic policies that soured relations with Britain and Russia; (9) the balance of power of ethnic groups within Austria-Hungary; (10) the impact of the rapid and total mobilization of the Russian army due to the absence of a plan for partial mobilization; (11) the role and interests of military elements in the German government and their influence on foreign policy; (12) the personality of Kaiser Wilhelm of Germany (the propensity he had for war at that time, or the qualities of mind and personality which allowed him to be manipulated by others into declaring war).

A *history* of the First World War will focus on all of these, plus many, many others from among the universe of factors possibly relevant to understanding why the war occurred. But a *theory* of the war focuses on just those few factors whose relations to each other can be shown to have been decisive in producing the particular outcome (war). Theory not only minimizes the number of factors to be treated as analytically relevant to an outcome, but, since it requires that we show how the factors are related to one another, it ensures that the

selection of factors is non-random. Let's say that, after devoting some thought and study to the causes of the First World War, you found your interest increasingly drawn to factors (5) and (12), above—Russian geopolitical ambitions and Kaiser Wilhelm's personality. Could you show that these factors not only helped to produce the war, but were also related to each other in some systematic and logical way? Later in this chapter we will look at a well-known theory of the causes of the First World War, discuss the variables on which the theorist chose to focus, the reason he chose them, and his argument about how they were related to each other and to the outbreak of war.

We have endeavoured to define 'theory' by contrasting it with what might be described as 'non-theory'. We can further define the term by relating it to two others that tend to be used interchangeably with it: 'proposition' and 'hypothesis'. Together, the three terms can be seen as representing different stages in the development of an idea. The first stage would be a **proposition**: a hunch or guess that two or more variables are related. When put forward for investigation—stated in a way that enables us to determine whether it is right or wrong—a proposition becomes a hypothesis. So an idea may start as a proposition, a statement that identifies key factors that are thought to be related in some way; it might then be developed so that the factors and their relations to each other are more precisely defined. This more specific statement is what we call a 'hypothesis'. Though a hypothesis is stated in a way that enables us to evaluate, analyse, or investigate a proposition, it is still a provisional and untested idea. It is only when it has withstood repeated tests and has been found to have considerable explanatory power that it becomes a theory.

There are all sorts of theories, and these can be distinguished in different ways. For instance, types of theories are distinguished by their scope and level of generalizability ('grand' versus middle-range theory), and analytic process (inductive versus deductive theory), and by the nature of the questions they address (empirical versus normative).

Consider the distinction that the sociologist, Robert Merton, defined between what he called '**grand theory**' and '**theories of the middle-range**' (Merton 1968). The definition of these types of theories, and the difference between them are based on their *scope and level of generalizability*. Grand theory is what he characterized as 'all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behaviour, social organization, and social change' (1968: 39). Merton argued that grand and abstract theories of society provided no basis for an empirical social science. He therefore called on theorists to apply themselves to the development of what he called 'theories of the middle range': theories that attempted to understand and explain a limited aspect of social life, a more restricted domain or set of social phenomena. These explanations could then be verified through empirical research and then perhaps systematized into theoretical systems of broader scope and content.

Theories also are disintinguished according to *the analytic process that links theory and research*, in particular whether theory guides research (deduction) or whether theory is an outcome of research (induction). We discussed these different processes in Chapter 2. Deduction moves from broader generalizations and theories to specific observations. Using this process, we start either with a theory that has already been confirmed or with a logical argument, and then we draw out the meaning or implications this has for explaining

some particular case or phenomena. So, in deductive theory, a hypothesis is deduced from current theory, which is then subjected to empirical scrutiny. Induction, you will recall, is a process of reasoning from particular facts to a general conclusion. We begin with particular observations or cases and then develop generalizations about them. So while deductive theory guides research, inductive theory is the outcome of research.

Another analytic process that links theory and research is what has been called '**grounded theory**' (Glaser and Strauss 1967; Corbin and Strauss 1990). Grounded theory is an inductive research strategy. The researcher starts by collecting data, and allow concepts and categories to emerge from them. Hypotheses are then developed through the interaction of theoretical elements and data. Consequently, theory is produced through, and *grounded* in, data. What is most emphasized in grounded theory is that it is explicitly an emergent process: the aim is to discover the theory implicit in the data, to allow theory to emerge from the data as opposed to forcing it into preconceived frameworks.

Finally, we can distinguish different types of theory according to the nature of the questions they address (empirical versus normative). We have previously discussed (and raised some objections to) the distinction between empirical theory and normative theory (Chapters 1 and 3). Empirical theory is concerned with questions that can be answered with empirical data (data gathered through observations of the world around us). **Normative theory** is concerned with questions about what is right and wrong, desirable or undesirable, just or unjust in society. We will have more to say about answering normative questions later on in this chapter.

So, having reviewed different types of theory, we can now sum up how the terms 'proposition', 'hypothesis', and 'theory' differ from, and are related to, one another. A statement positing that two or more variables are related is a proposition: a provisional idea that merits evaluation. In order for us to evaluate its worth, its constituent terms need to be defined very specifically. Once you have done this, you have a hypothesis: a tentative answer or argument you wish to develop (investigate, demonstrate) in response to your research question. A theory identifies a small number of variables that must be taken into consideration in addressing the question, and how they are related both to each other and to the outcome that is being addressed.

Where do hypotheses come from?

A hypothesis is developed through a critical analysis of the strengths and weaknesses of existing theory and previous research relevant to a given question or problem. As we discussed in Chapter 4, your analysis of existing theory and previous research relevant to your topic is presented in what is conventionally referred to as a 'literature review'.

Any issue relating to the political world is likely to be the subject of conflicting views about how to define it; where, when, and why it originated; how it developed or evolved; what its impact is; what its likely future trajectory will be; and what decision-makers or specialists can or should do with respect to it. In the process of sorting through the various opinions and judgements about these issues, you gain an understanding of the different positions and the current 'state of play' regarding a given topic. Who else has asked and answered your question? What puzzles or problems have they left unsolved? Are there important issues or dimensions of a phenomenon that they have not addressed? Are there

claims for which there is little systematic evidence; or relevant evidence or alternative data sources that have been insufficiently considered? Are there logical inconsistencies or errors in how variables or a relationship between two variables are specified, so as to cast doubt on the merit of the conclusions? Are concepts defined wrongly, ambiguously, too narrowly or too broadly, or in a way that obscures crucial dimensions of the phenomenon investigated? Do arguments and theories tend to focus on examples or cases drawn from only one part of the world? Do they fail to address or control for confounding factors or alternative explanation? Is the time frame too short or artificial; or the spatial domain too limited or artificial? What is missing, irrelevant, defective, unwarranted, or ambiguous? What should be further highlighted, underlined, expanded, extended, improved, built upon, or continued (e.g. through application to other cases)? In sum, what requires further investigation or re-thinking: what would provide a better understanding of the issue; make visible some dimension that brings added strength to one or another of the positions in a debate; produce a fruitful synthesis of different perspectives; or reveal erroneous assumptions or logic in current understandings?

By considering these questions, you start to identify good reasons for why you think one position is better, more persuasive, accurate, or comprehensive than another. Perhaps you start to develop a list of reasons for why you think one thing is true and another thing false and, eventually, for why one perspective is better than the others. In this way you begin to develop clear, focused, and logically sound reasons for thinking what you think. Your literature review retells this process. It identifies, and develops an argument about, what you consider to be weaknesses in the literature that need to be addressed, or strengths that have been insufficiently exploited. It is through this retelling of your critical analysis of the literature that your own argument or answer gradually emerges.

Your argument may initially take the form of a proposition that identifies key factors that you think are related in some way. To develop your argument into a hypothesis you will need to define the factors and their relations to each other more precisely. It is only by developing a hypothesis in this way that you will be able to investigate your argument.

We have been discussing how you arrive at a hypothesis through a critical analysis of the strengths and weaknesses of existing theory and previous research relevant to your question. It is worth pointing out that existing theories can help you develop your own hypothesis in a number of ways. Recall that, what we call a 'theory' is a hypothesis that has withstood repeated tests and has been found to have considerable explanatory power. Consequently, researchers may choose to use an existing theory to make sense of a particular case if they can show, for instance, that by *applying it to a case to which it has not previously been applied*, they are contributing to knowledge by further testing or refining it. For example, in her study of risk-taking in US foreign policy decision-making, Rose McDermott (2001) uses a hypothesis drawn from prospect theory to investigate a set of cases to which the theory has not previously been applied. Prospect theory maintains that 'risk-taking behaviour is based, not on the predispositions of a particular leader, but evolves out of a cognitive response to a situation that constrains the way options are interpreted and choice is made' (McDermott 2001: 4). More specifically, it proposes that individuals tend to be risk averse when they are in a good position (in the 'domain of gains'), and more likely to take risks when they are in a bad situation (in the 'domain of losses'). McDermott uses this proposition as a hypothesis to explain the decisions reached in four historical cases of US foreign policy decision-making

involving decisions taken by Jimmy Carter and Dwight Eisenhower in the domains of both losses and of gains.

A second way that theory can help you develop a hypothesis is when you *use variables already identified in the literature but combine them in a new way* to make sense of a particular case. This is what Jack Levy (1990–1) does to develop a hypothesis about the causes of the First World War. He uses variables proposed in other, more general theories of war—domestic political structure, misperception, and alliance structure—and then combines these variables in an original way which, along with novel causes distinctive to his case, enables him to produce an original explanation of the war.

Next, we provide an illustration of how one researcher goes about developing a hypothesis through a review of existing theoretical perspectives.

Illustration: how an analysis of existing theoretical perspectives provides the basis for a hypothesis

The process of reading and analysing literatures related to your research question can provide the basis for a hypothesis that escapes the weaknesses of existing studies and builds upon their strengths. Let's consider how a researcher's 'literature review' leads to the development of a hypothesis about an important area of inquiry in Politics and International Relations.

In *States, Scarcity, and Civil Strife in the Developing World* (2006), Colin Kahl is interested in exploring the role that environmental pressures play in generating civil conflict in developing countries. As Kahl observes, 'civil strife in the developing world represents perhaps the greatest international security challenge of the early twenty-first century'. In developing his argument about these conflicts, Kahl engages in a critical analysis of the large literature which focuses on the role in these conflicts of rapid population growth, environmental degradation, and competition over natural resources (2006: 1–2). His first step is to sort through and organize the various arguments and judgements about the relationship between environmental pressures and civil conflict that it contains. As we have previously discussed, this inventory and organization of existing knowledge on a topic is what enables a researcher to begin the process of developing and making a case for her own point of view. After some consideration, Kahl determines that he can bring organization to the large literature on resource-related conflicts by dividing the various arguments it contains into two broad perspectives: those which link conflict to *scarcity and grievance*, and those arguing that conflict is driven by *abundance and greed*. To analyse the strengths and weaknesses of these perspectives, he selects those hypotheses from within each perspective that appear to be most robust and influential. From the perspective that links conflict to resource scarcity, he selects the 'deprivation hypothesis' and the 'state failure hypothesis'; and from the perspective that links conflict to resource abundance he chooses the 'honey pot hypothesis' and the 'resource curse hypothesis'. He then analyses the strengths and weaknesses of each of these and, with this done, he is then ready to state his own hypothesis.

Here is a brief summation of Kahl's analysis of the four hypotheses and of how the hypothesis he develops based on this analysis addresses their weaknesses and builds on their strengths.

Kahl first considers the two hypotheses that focus on the link between resource scarcity and conflict: the 'deprivation hypothesis' and the 'state failure hypothesis'. The 'deprivation hypothesis' maintains that population growth, environmental degradation, and poor distribution of natural resources create deprivation among the poorest elements of society and, in this way, increase the chances of civil strife. Population and environmental pressures contribute to falling wages, unemployment, and landlessness. This increases poverty and inequality, leads to frustration and grievances, and increases the risks of collective violence.

Kahl points to two key weaknesses with this hypothesis. First, it ignores the collective action problem. This 'problem' is one which scholars in our field have grappled with for decades. It concerns the situation in which the existence of common interests among individuals fails to provide an incentive to pursue collective political action on behalf of those interests. What researchers have observed is that, rather than working to promote a common interest, individuals are just as likely to let others do the work and to 'free-ride'. The 'problem', then, is in assuming, as much political research seems to do, that collective interests lead to collective action. However, this is something that cannot be assumed but must be demonstrated. Kahl points out that the 'deprivation hypothesis' doesn't take this problem into account; it doesn't provide an explanation of when and why individuals are willing and able to make the sacrifices (in land, wages, and time) necessary to participate in organized violence under conditions of resource scarcity. The second weakness Kahl finds with this hypothesis is that it assigns no role to the state. The state, by virtue of its control over resources within a territorial domain over which it exercises its authority must, Kahl reasons, play an important role in resource-related conflict.

The 'state failure hypothesis' escapes this weakness by placing the state at the centre of the story it tells about how, when, and why resource scarcity generates conflict. It argues that population and environmental pressures, and increased demands for costly investment, weakens the state by undermining its capacity and legitimacy, and thus, opens up 'political space' for violence. Kahl argues that this hypothesis assumes, wrongly, that environmental and demographic pressures lead to conflict only in countries with weak state governance. State strength, he points out, can also provide an avenue for resource-related conflict because it can enable state elites to politically exploit demographic and environmental pressures to engineer violence against social groups. By focusing on weak states only, the 'state failure' hypothesis therefore doesn't go far enough in elucidating the role of the state in resource-related conflicts.

Having analysed the two hypotheses which focus on *resource scarcity*, Kahl then considers hypotheses about how *resource abundance* leads to conflict. The general argument advanced by this perspective is that resource abundance leads to conflict by encouraging rebel groups to form and fight over valuable natural resources. This is what the 'honey pot hypothesis' argues. The problem with this hypothesis, Kahl argues, is that the 'greed-based logic' on which it depends is chiefly applicable to non-renewable mineral resources rather than to renewable ones. Non-renewable resources—oil, diamonds, copper—are those most likely 'to be implicated in violent conflicts' in situations in which there are abundant resources, because these 'tend to be much more valuable per unit of volume, geographically concentrated, and easily tradable than most renewable resources'. Kahl argues that, because the 'vast majority of examples of honey pot-driven conflicts revolve around oil,

precious metals, diamonds, and other valuable minerals' (2006: 18), the hypothesis applies only to countries with non-renewable mineral resources. The same weakness attaches to the 'resource curse hypothesis'. This is perhaps the best-known resource hypothesis about resource abundance and conflict. According to this hypothesis, state control of abundant supplies of valuable resources contributes to corrupt, authoritarian governments that become prime targets for rebellion. Like the 'honey pot hypothesis', resource curse accounts apply 'much more to countries dependent on the export of non-renewable resources than renewable resources' (2006: 19). Moreover, these accounts, Kahl argues, take insufficient account of the ways in which social and political factors affect the relationship between resource endowments and violence (2006: 21).

To sum up these points: Kahl shows that the most important hypotheses concerning resource-related conflict are weak in one or more of three ways. First, they pay insufficient attention to collective action problems. Second, they don't provide an understanding of the role of the state in resource-related conflicts. Either they assign no role to the state or, if they do, they fail either to take into account the possibility both of state weakness (failure) and state strength (exploitation), or to provide a sufficient understanding of the variables that affect state capacity. Third, they are applicable to countries with non-renewable mineral resources rather than those with chiefly renewable resources.

Kahl then produces a hypothesis of his own that addresses these shortcomings. His hypothesis identifies two variables (the state and 'groupness') that link resources to conflict. He argues that the state plays a crucial role in resource-related conflicts in developing countries, because resource pressures can either push a state towards failure or provide the state with opportunities for exploitation. Resource scarcity, by leading to state weakness (failure) can, in turn, lead to conflict; resource abundance, by contributing to state strength and exploitation, can generate rebellion and conflict. However this, for Kahl, doesn't tell the whole story, because it doesn't tell us when or whether environmental pressures will push a state towards failure or exploitation. This, he argues, depends on social and political factors which affect the relationship between resource endowments and the state. Kahl conceptualizes these factors in terms of 'groupness'—the ease with which countries divide into ethno-cultural, religious, or class factions—which helps to overcome collective action problems; and 'institutional inclusivity or exclusivity', which relates to the influence wielded by groups or organizations over the state. Kahl hypothesizes that the potential for conflict will increase where there's a high degree of groupness, and where exclusive institutions short-circuit cooperation and leave state elites free to instigate violence. Conversely, the potential for conflict decreases where there is a low degree of groupness and where institutional inclusivity facilitates societal cooperation in the face of a weakened state. 'Groupness' then is the variable in Kahl's hypothesis linking resources to conflict.

Figure 5.1 shows the main hypotheses in the existing literature, and the hypothesis that Kahl produces as a result of his analysis of their strengths and weaknesses. It is useful to reconstruct your hypothesis, in the same way as the hypotheses in Figure 5.1, using an arrow diagram to identify key theoretical variables and the direction of the hypothesized relationship between them.

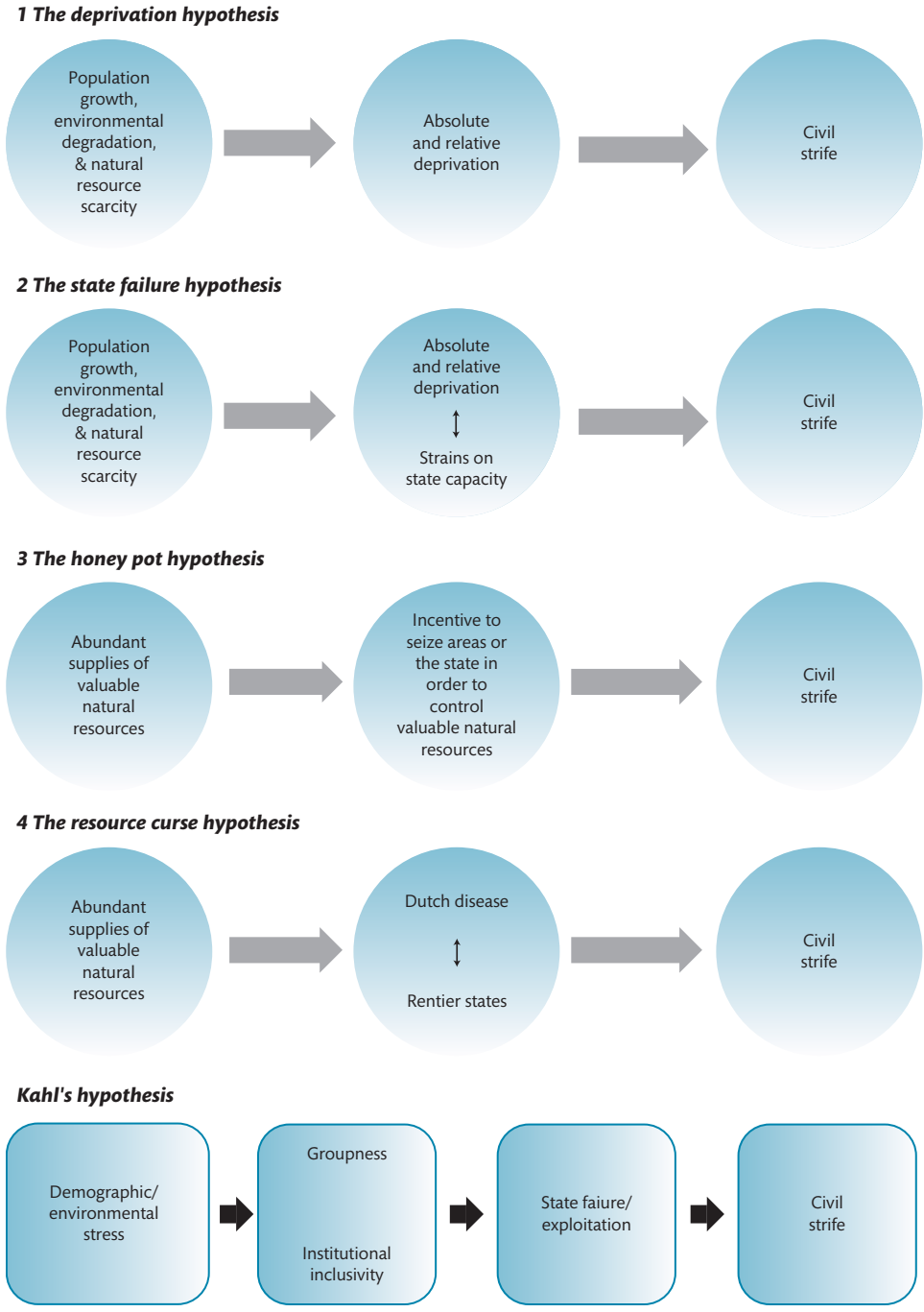


Figure 5.1 Hypotheses about how demographic and environmental pressures generate conflict

We have discussed how hypotheses emerge or develop through engagement with existing theory and/research. So when you state a hypothesis you need also to identify the source of your hypothesis—what, in theory or practice, gave rise to it. You will recall from Chapter 4 that we discussed the need to provide a rationale for your research question. You need to provide a rationale for your hypothesis as well. Was your hypothesis drawn, or developed deductively from a more basic theory. Is it based on assumptions of a general 'approach' (rational choice, institutional, Marxist, etc.), or a blend of aspects of two or more approaches? Or is your hypothesis developed inductively from a body of empirical evidence? Are you hypothesizing that the lessons drawn from another domain of experience can be usefully employed to address the domain relating to your research question?

All hypotheses are based on a set of expectations derived from existing theory and previous research. Whether the question you are addressing is descriptive, explanatory, predictive, prescriptive, or normative, there will be something in theory or previous research that provides the basis for a hypothesis about the way the world works or should work, about future developments or policy proposals. *The key components of some theory or explanatory approach or body of research are what lead you to expect the relation(s) your hypothesis states.* Together these components comprise what we refer to as a 'theoretical framework': the logic, story, or set of assumptions that connects your key factors or elements to each other. We said previously that theory ensures that contributions to knowledge are cumulative. But theory also furthers the research enterprise by providing patterns for the interpretation of data; by supplying 'frameworks within which concepts and variables acquire substantive significance', and by allowing us 'to interpret the *larger meaning* of our findings for ourselves and others' (Hoover and Donovan 2004: 37).

Let's consider now how hypotheses can be investigated. A hypothesis can be investigated in two ways. It can be *tested* with evidence (confirmatory research). Hypothesis testing uses logical or empirical evidence to evaluate existing theories. A hypothesis can also operate as a guide to a process of discovery (exploratory research). What we will call '*hypothesis-generating*' research produces findings that can be used in the development of theory.

Hypothesis testing begins by stating, on the basis of a theory, what we would expect to find, and then sees whether that expectation is fulfilled. Hypothesis-testing or confirmatory research is deductive in that it is driven by a particular hypothesis: the researcher has a specific, focused statement in mind and their objective is to prove or disprove that hypothesis. Hypothesis-generating research begins with a question and perhaps a basic proposition, examines a set of cases, and comes up with a more specific set of propositions. Hypothesis-generating or exploratory research is inductive, in that the researcher observes a phenomenon in order to generate questions or hypotheses for subsequent research. Both kinds of research are 'part of the overall scientific method' (Gerring 2001: 23). Both types of research must ultimately specify the relationship between variables. Both require one to make an argument. Both are part of a single, evolving, dynamic process of discovery and hypothesis formation which, by generating continuous interaction between theory and observation, moves the researcher from observation to theory (induction) and from theory back to observation (deduction). Figure 5.2 illustrates this process.

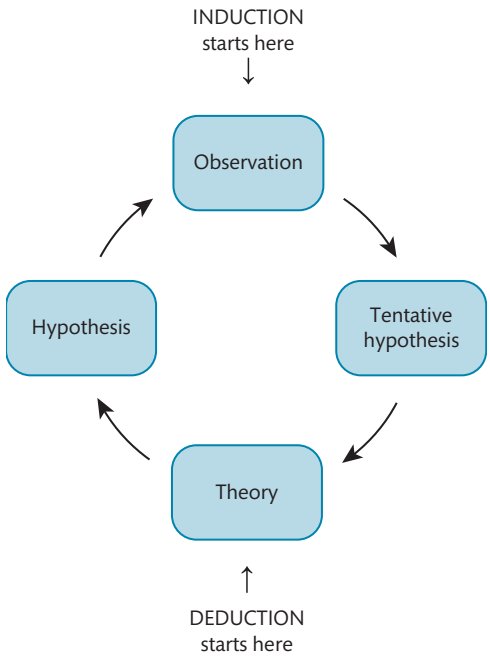


Figure 5.2 Induction and deduction

At the beginning of this section we stated that in order to contribute to knowledge in a field of study, an answer to a research question must meet three requirements. The first requirement is that the answer should be appropriate to the type of question that is being asked. The second is that the answer should make a contribution to knowledge in that field by relating the question and answer to existing research and theory. We have not yet discussed the third requirement, which is that it must be clearly and fully *specified*. We address this requirement in the next section.

Specifying your argument or answer: the nuts and bolts

We previously said that an idea may start as a statement that identifies key factors that are thought to be related in some way (a proposition), but that we cannot evaluate, analyse, or investigate this idea until the factors and their relations to each other are more precisely defined. This is what a hypothesis does. To develop a hypothesis you must clearly and fully *specify* its component factors and relations. You must not only specify its essential elements—the variables and their relationship to one another—you must also make sure you are using terms that mean what you want them to mean, and define them so that they can be understood by others in precisely the way you want them to be understood. This requires that you, first, develop a conceptualization, or conceptual definition, of the terms or variables you use. Second, you must develop an even more specific, operational, definition that identifies empirical referents for them.

The components of a hypothesis

A hypothesis consists of three elements: an *independent variable*, a *dependent variable*, and a statement about the *relationship* between the variables.

Variables

A variable is a concept or factor that can vary, change, or assume different values or characteristics. Variation in this factor contributes to variation in the outcome that we wish to understand. A factor that is not thought to vary in the context of the research and, hence, is not capable of contributing to variation in the outcome, is called a 'constant'. A variable assumes at least two different values or characteristics (e.g. high vs. low, more vs. less, present vs. absent). The values of a variable must be mutually exclusive: each observation must fit into one and only one category. Take a look at the examples in Box 5.4.

BOX 5.4 Variables and Values

Variables	Values
Resources	abundant/scarce
Interdependence	vulnerability/sensitivity
Domestic support for war	strong, neutral, weak
Form of government	democracy, dictatorship
Socio-economic and political changes	rapid/slow
Organization size	large/small
Democracy	stable/unstable
Economic development	low/moderate/high
Gender	male/female

In social science research, we deal mostly with two types of variables. The first type, an **independent variable**, is a factor thought to influence, affect, or cause variation in the second type of variable. It always comes before that other factor (the 'dependent variable') in time and space. We refer to this factor as 'independent' in the sense that, in the context of the research, we are not interested in investigating what brought it about. For instance, the variable 'environmental stress' in Colin Kohl's hypothesis, is an independent variable because, in the context of his research, we are not concerned with explaining how this stress came about. A **dependent variable** is the outcome that we wish to understand (conflict is the dependent variable in Kohl's hypothesis). We call it a 'dependent variable' because we hypothesize that it depends upon or is caused by variation in an independent variable. The relation between these two variables is often affected by an **intervening variable**. An intervening variable that affects the relationship between the independent and dependent variables by producing an interaction effect and acts as a 'moderator' variable. One that

transmits the effects of the independent variable to the dependent variable is called a 'mediating' variable. Without these mediating variables to act as a conduit, the independent variable would not affect the dependent variable. Kohl's hypothesis identifies 'the state' and 'groupness' as variables that intervene between environmental stress and conflict. Recall our discussion, in Chapter 2, on causal mechanisms and the notion that there are 'social mechanisms' that produce social outcomes. We related this issue to the question of how to provide explanations of macro-social phenomena with micro-foundations. Remember James Coleman's contention that macro phenomena have to be explained by the interaction of micro- and macro-levels ('Coleman's bathtub')? One of the functions that mediating variables might fulfil is to show how micro and macro phenomena are linked to each other. Box 5.5 summarizes the role of independent, dependent, and intervening variables.

We can illustrate variables, their values, and how they provide the nuts and bolts of an argument with further reference to the argument developed by Colin Kahl that we just discussed. The question Kahl is addressing is: what effects do environmental pressures have on conflict within developing countries? His dependent variable (what he wants to explain) is the likelihood of conflict. This variable can assume two values: *increased* likelihood or *decreased* likelihood. His independent variable (what he thinks affects this outcome) is demographic and environmental stress. These pressures can vary between two values: *high* and *low*. He identifies an important intervening variable that enables the independent variable to affect the dependent variable. This is the state: a mediating variable that links the effects of environmental pressures to the likelihood of conflict. This variable has two values: it can be *strong and exploitive*, or *weak and failing*. Which of these two values it assumes depends on other intervening variables having to do with the social and political institutions of a given country: 'groupness' (values: *high*/*low*) and institutional inclusivity (*high*/*low*). He argues that these variables affect whether environmental pressures push a state towards failure or exploitation.

BOX 5.5 Variables

Variable: a factor or characteristic that can assume different values or characteristics

- A **dependent variable** is the variable in which you are primarily interested, and which you assume is dependent upon a second variable. Our dependent variable is some political outcome (events or process) we are endeavouring to explain.
- An **independent variable** is the variable thought to, directly or indirectly, cause or influence the dependent variable. Within this particular relationship it is independent because it does not depend on the other variable/variables. Our independent variables are the factors that bring about the outcome we wish to explain—the factors that are determinative of this political outcome. The value of the dependent variable is influenced by (or depends on) the value of the independent variable/variables.
- An **intervening variable** influences the relation between two others and thus produces an interaction effect *('moderator' variable)* or transmits the effects of the independent variable to the dependent variable (a *'mediating' variable*).

Relationships

All types of research questions are concerned with connections or relations between two or more variables. A descriptive question might ask *what* relationship exists between gender and voting preference, or between educational achievement and political participation; an explanatory question would ask *why* that relationship exists. A predictive question would ask *how that relationship might be expected to change* in the future; a prescriptive question might be concerned with *what steps might be taken to strengthen a relationship*. While we don't think of normative political theory as requiring a dependent and independent variable, normative questions nonetheless are concerned with connections and relations, and with *whether they are just or desirable*.

There are two types of relationships in political research, two ways that variables can be connected or related to each other. In practice, these two types tend to converge.

The first relationship is one of **association**. In this relationship, a variable, for example 'unemployment', is in some way associated with, related to, or linked with another variable, for example 'inflation'. This term is roughly synonymous with **correlation** (two things vary together in a linear fashion) and with **co-variance** (the alleged cause varies with the supposed effect). All these terms refer to a relation between variables such that changes in one variable *occur together* with changes in the other. A relationship of association can be positive or negative. Two variables are *positively related* when they change in the same direction: when the occurrence of high values on one variable are associated with the occurrence of high values on the other, and low values on one are associated with low values on the other. For instance, a positive correlation between inflation and unemployment means that when inflation is high, unemployment also tends to be high; and when inflation is low, unemployment also tends to be low. When two variables are *negatively related*, the values of the variables change in opposite directions, so that when one variable (inflation) decreases, the other (employment) increases.

The second type of relationship between two variables is **causality**. With this type of relationship, changes in one variable *bring about* changes in another. To establish that a causal relationship exists between two variables you must show that four conditions have been met. The first condition is that the hypothesized cause or independent variable is temporally prior to the effect, that is the dependent variable. In other words, the independent variable must precede the dependent variable in time. The second condition is that the two variables are correlated or co-vary. Of course, correlation does not establish that two variables are causally related. That is why we need the third condition: that a causal mechanism or process links the two variables. We discussed causal mechanisms in Chapter 2, and the desirability, more generally, of showing the 'micro-foundations' that connect variables—the intentions and choices of individuals, or of social and political units that have acquired the capacity to act collectively. You have to be able to tell a plausible story that you think probably connects the independent variable(s) with the outcome that you are trying to explain, often including an 'intervening variable' that gets us from the hypothesized cause (independent variable) to the effect (dependent variable). The fourth condition is that the correlation between the independent variable and the dependent variable is not spurious, meaning that you have ruled out the possibility the correlation that appears to exist between the two variables does so only because there is a variable, antecedent or causally prior to both the independent variable and the dependent variable, that is affecting both.

A well-known and puzzling finding about US Congressional electoral campaigns can illustrate how a third variable can render a seeming correlation spurious. In a study of the effects of campaign spending in congressional elections, Gary Jacobson (1978) found that increased spending by incumbents correlates with increased odds of failure. So, it might be assumed from this finding that *spending more* by incumbents somehow *causes* electoral failure. However, Jacobson found that there was a third variable that was antecedent to both spending and electoral failure: the nature of the electoral battle, that is, whether it was expected to be tough or easy. Incumbents who expect a tough fight spend more than incumbents who expect to win easily. If spending more is evidence of a tough fight, there is nothing puzzling at all about finding that increased spending appears to be correlated with electoral failure. The notion that somehow spending more by incumbents *causes* electoral failure is spurious: another variable, the toughness of the fight, explains both high spending and electoral failure.

In sum, when we assert that variation in independent variable X *causes* variation in dependent variable Y, we are making four assertions: (1) the change in X precedes the change in Y; (2) X and Y are correlated; (3) a causal mechanism or process can be identified that links the variables; and (4) the correlation between X and Y is not spurious or a coincidence.

We said that in practice the two types of relationship we've discussed—association and causality—tend to converge. To explain why they do, we need to explain that there are two different notions of causality that social scientists employ: deterministic causation and probabilistic causation. A *deterministic* causal relation states that 'if (X) then *always/invariably* (Y)'. However, we generally tend to see human behaviour as not determined, but constrained; consequently Politics and International Relations research usually employs a *probabilistic* notion of causality rather than a deterministic one. A probabilistic causal relation states that 'if (X) then *maybe/sometimes/probably* (Y)'. 'Probabilistic' means that when the values that an independent variable takes on increase, this *usually results* in the values of the dependent variable increasing (or decreasing). This notion of cause focuses the efforts of researchers on finding factors that make an effect more likely: for example, the finding that the more educated a person is, the *more likely* they are to vote. The more robust we make our correlations—the more we seek to meet the conditions of causality discussed earlier—the more the relation of correlation between two variables converges with one of probabilistic causation. Though a robust correlation cannot establish *why* X and Y co-vary, knowing that X is usually associated with Y can make a significant contribution to our understanding of political outcomes.

Conceptualization and operationalization

In clearly and fully *specifying* our hypothesis we need to consider not only its key components and the relations among them: we need also to ensure that the terms we use mean what we want them to mean and are defined in a way that will enable others to understand them in precisely the way we want them to be understood. To do this, we must provide both a conceptualization or *conceptual definition* of them, as well as a definition that identifies empirical referents to them: an *operational definition*. Specifying our terms in this way makes it possible for our ideas to be assessed and for our research to enter into dialogue with and build on the work of others.

Concepts and concept formation: what are you talking about?

A concept is a term applied to a collection or class of things that are to be regarded alike, because they either share a common characteristic or feature (e.g. gender, type of government) or are instances of some general type of behaviour (e.g. war, participation). A concept provides a label for instances or characteristics that are commonly found together. When we attach a label or general term (concept) to observations or events, we can link separate observations or events together and form generalizations. We observe people voting and we refer to this as 'participation'. We observe that a legislator consistently takes certain positions on political issues and we call them a 'conservative'. Translating knowledge into the more general categories represented by these concepts enables knowledge to cumulate.

We use concepts as a shorthand to refer to much of the substantive content and issue base of political studies: elites, bureaucracy, legislatures, judiciaries, policy, ethnicity, pluralism, alienation, human rights, government, federalism, public opinion, elections, power, development, democracy, culture, legitimacy, charisma, hegemony, institutionalization, exploitation, authority, interests, class, corporatism, civil society, racism, terrorism, egalitarianism, dependency, consensus, welfare state, social justice.

But, many concepts used in our field are anything but clear. There is often a lack of consensus about which concepts are important, what they mean, and how they should be defined and measured. What one scholar means by 'ideology' may be quite different from what another means by it. There is also a lack of consensus about which concept should be employed to describe a given phenomenon or outcome. Researchers frequently use different concepts to understand the same situation. What one researcher calls a 'terrorist organization' another might call a 'social movement'; an event that one researcher calls a 'revolution' another might call a 'coup'. Social science is characterized by a continuous conflict not only over how we define concepts, but which ones we should use to comprehend a given reality. Consequently, we need to spell out with as much precision as possible what a concept that we use (e.g. 'ideology') means in the context of our research. The point is not just to define the terms we use, but to consider very carefully *what* terms we should use and *how* we should define them.

Concept formation

The term '*concept*' refers to a mental image (conception) that summarizes a collection of seemingly related observations and experiences. *Conceptualization*, or 'concept formation', is the process of selecting the term by which some collection of things should be known. As John Gerring points out, it concerns 'the most basic questions of social science: *What are we talking about*?' (2001: 35). Much thought has been devoted to the question of what distinguishes a *good* concept from one that is less good or less useful. Gerring identifies eight essential criteria for a good concept: familiarity, resonance, parsimony, coherence, differentiation, depth, theoretical utility, and field utility. Here is a list of these criteria and the question(s) that each raises and requires you to consider (Gerring 1999: 367):

1. **Familiarity:** how familiar is the concept (to a lay or academic audience)?
2. **Resonance:** does the chosen term ring (resonate)?
3. **Parsimony:** how short is the term and its list of defining attributes?

4. **Coherence:** how internally consistent (logically related) are the instances and attributes?
5. **Differentiation:** how differentiated are the instances and the attributes (from other most-similar concepts); how bounded, how operationalizable, is the concept?
6. **Depth:** how many accompanying properties are shared by the instances under definition?
7. **Theoretical utility:** how useful is the concept within a wider field of inferences?
8. **Field utility:** how useful is the concept within a field of related instances and attributes?

A good concept must mediate among these criteria, usually through a series of trade-offs. For instance, you decide that a good place to start the process of conceptualization is with a term that meets the criterion of *familiarity* (established usage). You find, however, that the term has little *theoretical utility* (usefulness for theory). Your next step will be to find a term that does a better job of mediating between these two criteria. Perhaps you come up with a new term for something that offers *resonance* (a 'ring'). You may discover, however, that this comes at too high a cost with respect to *familiarity*, *theoretical utility*, or other criteria. *Parsimony* (explaining the most with the least) must be balanced with *depth* (the number of features shared by the observations or instances we want to regard as alike). *Coherence* (the internal consistency of the instances and attributes) often needs to be considered in relation to *differentiation* (of the observations and instances from similar concepts).

We can illustrate the problems and trade-offs that researchers face in concept formation by considering how researchers have struggled to arrive at a satisfactory definition of a key term of political analysis: power. To do so, we provide a brief summary of the debate about how to conceptualize power. This is not intended as a full discussion of the debate, but only as an illustrative example of the problems researchers in our field confront when attempting to define key concepts.

Conceptualizing 'power': the search for theoretical utility

Researchers in our field generally agree that 'power' is a key concept in political inquiry, and that it is central to any understanding of politics. But what we mean when we use the term 'power' is anything but clear. Is 'power' a capability, an attribute, or a relation? A capacity to do something or control somebody, or an action? A property of actors or of relations? Is it necessarily conflictual, intentional, or effective, or all three? Who possesses and exercises it: individuals, groups, social institutions, and organizations?

Difficulties of this sort arise, in part, because of the problem of differentiating 'power' from similar or related phenomena, such as coercion, authority, influence, domination, force, inducement, persuasion, manipulation. But attempts to achieve a high degree of *differentiation* in the conceptualization of power often entail a loss of *depth*. The term 'power' must not only be differentiated from other similar concepts; it must also capture the many different ways that people can exercise power over each other (some of which cannot always be observed), and the very diverse sources of power: material resources like money, property, wealth, natural resources, information, military capabilities, good jobs; but also intangible factors like charisma, status, prestige, reputation, respect, honour, character.

Let's consider these difficulties in relation to two ways in which researchers have attempted to define 'power': in terms of relations between actors (power as control over actors, agendas, and preferences); and the outcome of those relations (power as control over events and outcomes).

A conceptualization of power as *control over actors* defines 'power' as the ability of one actor to impose their will on others. This fits our intuitive notion of power (*familiarity*); and it captures several dimensions of power (*depth*) by including control of actions, agendas, and preferences. In this conceptualization of 'power', to use Robert Dahl's oft-quoted formulation, power is the ability of A to get B to do something she would not otherwise do (Dahl 1957). One way this can be achieved is through controlling *agendas*—structuring or determining what issues are discussed, preventing 'the public airing of policy conflicts' (Bachrach and Baratz 1962: 949); keeping potential issues out of politics, 'whether through the operation of social forces and institutional practises or through individuals' decisions' (Lukes 1974: 24). The importance of controlling agendas is illustrated by the struggle of countries in the global South to gain some control over the agenda relating to North–South issues; the agenda-setting power of media, non-governmental organizations, or global social movements, like the anti-globalization movement. 'Control over actors' can also be achieved through controlling *preferences*—through influencing what other actors desire. The ability to get what you want, not through the 'hard power' of military and economic strength, but through attraction, is what Joseph Nye (1990, 2002, 2004) has called 'soft power'. Controlling preferences means the ability to generate consensus, through the manipulation of public opinion or propaganda (disinformation), through censorship (withholding information), or political correctness (thought control).

Controlling actors through control over actions, agendas, or preferences encompasses much of what we mean when we use the term 'power'. However, this conceptualization makes it difficult to distinguish between interactions in which one actor exercises power over another and those involving exchange relations. How do we know whether the outcome of an interaction between two actors (e.g. a negotiation) occurred as a result of the exercise of power by one actor over another, or because both sides had something to gain from the outcome? If the UK agrees to the demand by the US to impose taxes on offshore activities, does this show that the US exercised power over the UK with respect to this issue, or that some sort of exchange occurred in which both benefited? Can we interpret the outcome, more generally, as demonstrating something about current power relations in the international system, or something about the nature and content of exchange relations? A 'control over actors' definition of power often makes it difficult to distinguish one from the other.

But there is another problem with this conception of 'power'. It is tautological: you are defining power in terms of the outcome. Let's say that we think that some action by Actor B can be explained as a result of the power that Actor A was able to exercise over Actor B. We explain that B did what A wanted them to do because A had the power. But we can only know that A had this power because B *did do* what A wanted them to do. We only know that A has power over B by the result. 'Power' as a concept is only interesting if it can be used to explain events or outcomes; and a conceptualization of 'power' that is tautological is not *theoretically useful*.

Consider the second conception of power: *control over events and outcomes*. By focusing not on relations between actors, but on the outcome of those relations, this conception

allows us to account for interactions between actors in which both actors benefit (exchange), though they may gain disproportionately. In a conception of power as control over the outcomes, we can not only account for shared power or exchange, but also for situations in which relatively weak countries exercise power through sharing control over mutually consequential outcomes. For example, during the Cold War, the small, relatively powerless Middle East states exercised power through their shared control over stability in the region. They could threaten to act in destabilizing ways: to escalate the Arab–Israeli conflict and cause a widescale confrontation in the region, or to lose militarily. But though this conception of power offers important advantages over the 'control over actors' conception, it doesn't escape the problem of tautology. Like the 'control over actors' conception, it defines power by the outcome; so there is no objective way of determining who has the power until we see how things turn out.

A more general criticism of both conceptions is that, in general, most of what happens in politics is not attributable to the direct exercise of power by actors, over each other or over outcomes. Power is usually exercised in a far less direct way. This is what the notion of *structural power* highlights: aspects of power which operate through indirect institutional, unintended, or impersonally created effects. Structural power operates through a network of (historically constituted) relationships that benefit some and disadvantage others (Guzzini 1993). Susan Strange (1994) identified four key structures through which power is exercised: the *security structure*, which provides protection to human beings; the *production structure*, the primary means of creating value and wealth in a society; the *finance structure*, which determines who has access to money, how, and on what terms; and the *knowledge structure*, which 'determines what knowledge is discovered, how it is stored, and who communicates it by what means to whom and on what terms' (1994: 121). When women of childbearing age in poor countries work at factory jobs that expose them to potentially dangerous chemicals, it is a result of the exercise of power, not by employers, but of production and financial structures that give poor women in these countries few other ways to survive. The US is able to indirectly influence other states' behaviour by controlling the monetary structures within which they must operate.

We have touched on only a few of the challenging and vastly complicated issues and considerations involved in the attempt to conceptualize 'power'. Hopefully, the examples have sufficed to alert you to the sort of ambiguity that attaches to terms that are commonly used in our field and the consequent need to provide a carefully considered conceptual definition of those that you use.

Conceptualizing 'democracy': mediating between coherence and differentiation

We achieve broader knowledge by generalizing; and we generalize by applying our conceptualizations to a wider range of observations or experiences. However, using existing concepts to cover instances that lie 'outside their normal range of use' (Gerring 1999: 360), either in application to new cases or as a result of changes over time *within* cases—captured by Giovanni Sartori's (1970, 1984) notions of *conceptual travelling* (the application of concepts to new cases)—can create problems of *conceptual stretching* (Collier and Mahon 1993: 845)—the distortion that occurs when a concept does not fit the new case. We can illustrate

the problem of 'conceptual stretching' and efforts to prevent it, by considering attempts to deal conceptually with the changing nature of political democracy over the past decades.

The great diversity of post-authoritarian political regimes established in Latin America, Africa, Asia, and the former communist world share important attributes of democracy. But they also differ in significant ways both from each other and from the democracies in advanced industrial countries.

The term 'democracy' has been applied to those political systems that meet a 'procedural minimum' definition of 'democracy': fully contested elections, full suffrage, the absence of massive fraud, and effective guarantees of civil liberties, including freedom of speech, assembly, and association (Collier and Levitsky 1997: 434). Those political systems that have this procedural minimum, plus some further differentiating attribute, provided the basis for defining the 'classical' sub-types of democracy as, for instance, 'parliamentary democracy', 'multiparty democracy', and 'federal democracy'. In order to capture the diverse forms of democracy that have emerged in recent decades, researchers have defined new sub-types: 'authoritarian democracy', 'neopatrimonial democracy', 'military-dominated democracy', 'limited', 'oligarchical', 'controlled', 'restrictive', 'electoral', 'illiberal', 'guarded', 'protected', and 'tutelary' democracy. The goal was to generalize our understanding of democracy to a wider set of cases. But because these 'sub-types' are cases that do not meet the definition of the root term, that do not represent *full* instances of the term, 'democracy', it could be argued that they really refer 'to a *different* set of cases' than those that we call democracies (Collier and Levitsky 1997: 430, 438).

Rather than referring to forms of democracy which are incomplete, that is, which are missing attributes included in the procedural minimum definition, as different *types* of democracy, some analysts consider it more useful to refer to political systems as exhibiting different *degrees* of democracy (Collier and Levitsky 1997: 440).

This approach owes much to Robert Dahl's influential study, *Polyarchy* (1971). In this study, Dahl uses the term 'democracy' to refer to a political *ideal*. The term 'polyarchy' refers to *actual* political systems. Polyarchies are systems of representative government that 'to a relatively high degree' are characterized by (1) equal chances of all members of the community determining the agenda for, as well as the outcome of, processes of political decision-making; and (2) the effective accountability of the elected officials during the time between elections (1971: 84). The notion that actual political systems can be characterized in terms of degrees or gradations of democracy has been widely accepted. Some analysts argue that 'democracy is *always* a matter of degree' (Bollen and Jackman 1989: 612; emphasis added; see also Dahl 1971, 1989; Bollen 1980, 1993; Coppedge and Reinicke 1990; and Bollen and Paxton 2000).

But others argue that this approach to conceptualizing democracy is fundamentally flawed. 'Political systems are *systems*', Giovanni Sartori argues. They are 'bounded wholes characterized by constitutive mechanisms and principles that are either present (albeit imperfectly) or absent (albeit imperfectly)' (1987: 184; emphasis added). Regimes 'cannot be half-democratic' (Alvarez et al. 1996: 21). The distinction between democracy and non-democracy should be treated, therefore, as exhaustive and mutually exclusive categories. The question of whether to conceptualize 'democracy' as a dichotomous variable (democracy versus non-democracy) or in terms of degrees continues to be much debated.

As the discussion of 'power' and 'democracy' shows, we must not just define our terms, but think carefully about the implications of how we use them and the trade-offs that must be made among a variety of considerations that enter into the formulation of good as a concept.

Operational definitions: how will you know it when you see it?

Even when there is a consensus about the importance of a particular concept and how it should be defined, researchers often will *dis*agree about whether, where, or to what degree it is occurring. For instance, beginning in the 1970s, International Relations scholars became increasingly preoccupied with the growing interdependence among states. Researchers have generally agreed on what the term 'interdependence' means; but they do not always agree about how significant it is, whether it is increasing or decreasing, regional or universal, new or old. Even though we may agree on the meaning of a term, unless we provide an 'operational' definition of it—define it in a way that will enable us to determine its presence or absence, strength, and extent—we may not know it when we see it. So, in addition to providing a conceptual definition of a factor or variable, we need an operational definition, as well. **Operationalization** means defining a variable, so that we know it when we see it.

Suppose you ask: 'Is Yemen a failed state?' You adopt a definition of 'failed state' used by other researchers: 'a failed state' is a state experiencing high state vulnerability or risk of violence. You now have a conceptual definition of 'failed state'. But how do you know 'state vulnerability' when you see it? You cannot leave it to your subjective sense of what constitutes vulnerability on the part of a state. You need to decide what will enable you to judge the degree of vulnerability of a state. You might decide that 'state vulnerability' can be operationally defined as a state in which there exists (1) humanitarian emergencies as a result of massive movement of refugees or internally displaced persons; (2) sharp and/or severe economic decline; (3) a progressive deterioration of public services; (4) the intervention of other states or external political actors. You can then go and get evidence and data on each of these indicators of 'state vulnerability' in order to reach your conclusions about whether or not Pakistan is a failed state.

We will not be content to know your conclusions: we need to know how you reached them. You must make explicit the set of actions (operations) that will lead another to achieve the same result when repeated. Other researchers should be able to follow your procedures and come up with the same findings. An operational definition provides travel instructions: sets of actions (operations) which will reveal to other researchers how you arrived at your findings. It specifies what empirical *indicators* or data will enable us to determine the presence or absence of a variable, or measure its extent. To operationalize the term 'speeding', we put a number on it: 70 miles per hour. This becomes a proxy for 'speeding'. It also suggests the data that will enable us to determine when 'speeding' has occurred: a speedometer reading. Let's say you hypothesize that the higher someone's social class, the more likely they are to be conservative. You might choose to use different income levels to operationalize 'social class': less than $40,000 is 'lower' class, $40,000 to $79,999 is 'middle class', and $80,000 and over is 'upper class'; and to operationalize 'conservative' on the basis of a series of questions about political issues. Perhaps you are interested in whether there is a difference in the 'personal adjustment' in the military of men and women. How do you know 'personal adjustment' when you see it? You decide that those who—in answer to the question 'How have you adjusted to being in the army?'—responded that in general they

were in good spirits liked being in the army, and that they were satisfied with their army jobs and status, and can be classified as well-adjusted personally.

Illustration: Lenin's explanation of the First World War

Let's review the components of a hypothesis by considering how they combine to produce a well-known theory of the causes of the First World War.

During the First World War, a book by Vladimir Ilyich Lenin (1870–1924) was published, entitled *Imperialism, the Highest Stage of Capitalism* (1916). The book was inspired by a concern to understand the causes of the war that had engulfed Europe and was escalating into a global conflagration. In the process of developing an explanation of the war, Lenin also developed an argument about capitalism, imperialism, and revolution.

Lenin explains the First World War as a consequence of intense imperialist rivalry among the major European powers. The source of his hypothesis was Marx's notion that the contradictions of capitalism would lead inevitably to revolutionary crisis and open the way to a transition to socialism. Marx had pointed to a number of contradictions of capitalist development:

1. Technological change under the direction of capitalism creates unemployment (a 'reserve army of labour'). This holds wages in check and depresses the consuming power of the working class.
2. Unemployment is an integral and essential element of the system: without it, capitalists would lose control over the labour market.
3. As capital accumulates, the rate of profit falls. This causes capitalists to interrupt the accumulation process, which produces economic crises and depressions.
4. The power of the working class to consume doesn't keep pace with the growth of production potential. As a result, production potential remains underutilized, and in place of actual economic expansion, stagnation and contraction eventually sets in.

Figure 5.3 shows how Lenin derived a hypothesis from Marx's theoretical work. Lenin observed that capitalist nations had avoided the crisis that Marx had predicted through overseas imperialism. He argued that:

1. Finance capital responds to the falling rate of profit by combining into larger economic units—cartels, monopolies—in order to gain market power.
2. Finance capitalists invest capital abroad in less developed regions to get a higher rate of return.
3. The flag follows finance capital because government is needed to protect investment, to set up exclusive spheres of investment. As capital accumulates and profit rates fall, capitalist economies in competition with one another embark on an effort to divide up the world in order to secure external markets, investment outlets, and sources of food and raw materials.
4. Once the entire world is split up, a re-division of territory can only come about through war. In the process of war, socialist revolutions occur abroad.

The imperialist solution to the problem of overproduction/underconsumption was to secure external markets and investment outlets abroad. This prevented socialist revolution

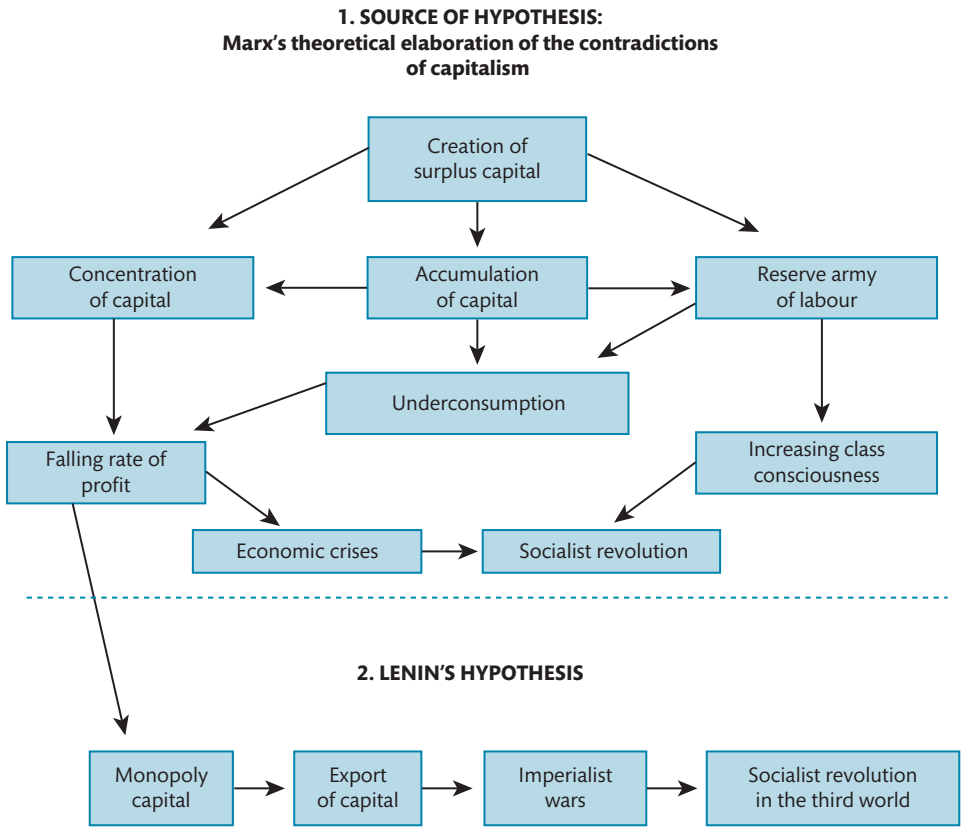


Figure 5.3 The Source of Lenin's hypothesis about the causes of the First World War

Source: Inspired by Gurley 1975: 54. Our thanks to David Lake, who used a variation of it in an undergraduate seminar at UCLA.

at home (and, at the same time, made socialist revolution much more likely to occur in the underdeveloped regions due to the hyper-exploitation of workers there). Lenin's hypothesis was that the First World War was the outcome of efforts by the imperialist powers to territorially re-divide the world among themselves. The temporal domain to which he intended this hypothesis to apply was 'the era of monopoly capitalism'.

We can identify the specific component of the hypothesis as follows. The dependent variable is war in the era of monopoly capitalism; the hypothesized independent variable is economic stagnation. Lenin connects the independent and dependent variables by introducing two intervening variables. The first is the imperialist expansion that ensues when economic stagnation leads to pressures for the acquisition of markets and investment outlets abroad. The second intervening variable is the availability of territory, which decreases as more and more capitalist countries pursue imperialist expansion in order to offset the economic stagnation generated by contradictions of capitalist production. There is a positive relationship between economic stagnation (the independent variable) and expansion (an intervening variable): as one increases, so, too, does the other. There is a negative relationship between the two intervening variables: as expansionism increases, the availability of territory decreases. There is also a negative relationship between the availability of

territory and the potential for war: a decrease in the availability of territory increases the likelihood of war. Figure 5.4 provides a diagram of these variables and how they are related.

After formulating this hypothesis, Lenin investigated it by looking for evidence that would support it or prove it wrong. He presented data and evidence for each of his key variables, and relevant conclusions from leading scholars and economists, both Marxist and non-Marxist, including data on economic stagnation in the major belligerent countries, policies of imperialist expansion on the part of those states, and the decrease in the amount of available desirable territory into which these states might expand.

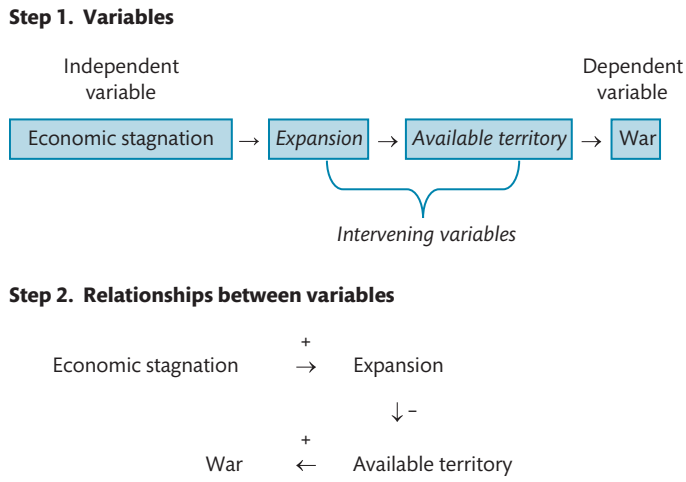


Figure 5.4 Lenin's hypothesis about the causes of the First World War: variables and relationships

Answering normative questions

It is sometimes noted that normative political theorists tend not to be as explicit about their methods as are researchers concerned with empirical questions—that they are often 'silent on questions of method and approach' (Leopold and Stears 2008: 1). David Leopold and Marc Stears observe that, in fact, 'the vast majority of students beginning advanced research in political theory in the United States and Britain' embark on their studies 'without any significant training in, and reflection on, the research methods that they will have to employ if they are to produce high quality work of their own' (2008: 2).

We have said that the research process we outline in this chapter is relevant to all types of research. Here we want to elaborate this point in relation to research that addresses normative questions. In our view, answers to normative and empirical questions can and should be developed in broadly similar ways.

Empirical and normative research

In previous chapters (Chapters 1 and 3), we discussed the distinction that has been institutionalized in political research between empirical and normative research. Empirical research is thought to concern questions about the 'real world', questions that can be answered with empirical data (data gathered through observations of the world around us); while

normative research is seen as concerned with questions about what is best, just, right, or preferable in society — questions which are best answered through forms of argumentative discourse. But, as we have suggested in previous chapters, the distinction defined between empirical and normative research is overdrawn. Among other things, this has likely helped to obscure the extent to which researchers use generally similar methods to address both.

The empirical/normative distinction is predicated on the assumption that it is possible to wholly separate questions about what *is* (empirical questions) from questions about what *should* or *ought to be* (normative questions). However, as we pointed out in Chapter 3, many people engaged in political research have challenged this assumption, and this includes many normative political theorists as well. In fact, there is an ongoing debate among political theorists concerning the relationship between normative political theory and empirical inquiry. The debate concerns what many perceive to be the gulf between 'ideal' foundational philosophical reflections on, for instance, rights, liberty, equality, or justice, and the real-world conditions and institutions in which problems relating to these issues arise and must be resolved. One position maintains that the point of normative political 'theory' is to arrive at a coherent *ideal*. The articulation of an ideal contributes to our practical understanding because it allows us to measure the distance between it (the ideal) and the real world. The other position argues that, since the ideal is constrained by what is possible, we need to integrate theory with a consideration of its implications and likely outcomes for the political world in which real citizens live. According to this view, the task of the political theorist is to investigate both political principles and the underlying 'facts' of political and social life that underpin them (Miller 2008). The aim of normative political theory is the clarification of underlying basic principles or arguments. But normative arguments that propose or justify one value system over another will lack relevance if they make no reference to related empirical and practical issues.

We can better understand the two positions by considering them in relation to theories of social justice. Theorists on one side of the debate argue that the fundamental principles of justice are logically independent of issues of feasibility and questions about human nature (e.g. Mason 2004; Cohen 2008); that only by removing itself from the world of empirical observation and the particularities of time and of place can normative political theory perform its proper task of identifying the actual meaning of 'justice', and of outlining the ways in which our world could be made to fit its demands (Stears 2005: 326). An example of what might be called 'ideal' theorizing is John Rawls' *A Theory of Justice* (1971). In this highly influential work, John Rawls presents what he identifies as an 'ideal' theory of justice. This theory is 'ideal' in that it makes two idealizing assumptions: that (1) all comply with the principles of justice; and that (2) social conditions are such as to enable citizens and societies to abide by principles of political cooperation. Rawls says that, by abstracting from the world as it actually is (e.g. ignoring the possibility of law-breaking or conditions that overwhelm the capacity of actors for moral reasoning), ideal theory allows us to find ideal principles of justice and to develop an understanding of how to reform our non-ideal world in line with these principles.

Theorists on the other side of the debate about normative theory and empirical inquiry argue that research on social justice has become detached from the real world; that theorists who focus on the question of what an ideally just society would look like neglect the empirical realities of real societies and the implications of their theorizing for pressing practical issues of social justice confronted by societies here and now. They argue that constructing

a theory that can help resolve problems of social justice requires consideration of the non-ideal conditions and social institutions faced by actual political actors (e.g. Dunn 1990; Carens 2000; Farrelly 2007; Sen 2009).

In our view, analytically rigorous theoretical endeavour need not be comprised, but rather might be enhanced considerably, by giving due consideration to the diverse complexities that arise in real, non-ideal societies. Depending on the question, political theorists might combine the use of philosophical methods of logic, categorization, or conceptual analysis with any of the methods of data collection that researchers use to address empirical questions.

Empirical and normative questions, answers, and methods

Reflecting the views concerning ideal theory and empirical inquiry discussed earlier is the tendency to think of normative questions as generally falling into one or the other of *two* basic *types*. The first type includes *questions in which principles, ideals, and concepts are the primary focus*. What *principles* ought to be adopted and enforced such that compliance with them will achieve social justice? What would an *ideally* just society look like? How is the *concept* of justice best understood? Theorists tend to address these sorts of questions by employing philosophical methods of logic, categorization, or conceptual analysis: by analysing the logical structure of a principle, systematizing the reasons that can validly be advanced for and against particular choices, or developing a conceptual framework. Theorists might engage in a form of argumentative discourse that, by drawing logical inferences and showing how the ideas presented were logically developed, convinces others that the conclusion reached is reasonable and plausible. Theorists might use any number of different methods of categorization to define the features that something has or doesn't have, by, for instance, comparing the similarity of an instance to a prototype. Theorists might break down or analyse concepts into their constituent parts using conceptual analysis: 'an attempt to provide an illuminating set of necessary and sufficient conditions for the (correct) application of a concept' (Audi 1983: 90), and a test of such sets using hypothetical examples and thought experiments.

The second type of questions normative political theorists ask are *questions that involve evaluating and criticizing substantive political, social, and economic practices and institutions*. Which institutions and practices are appropriate to implement the principles of distributive justice? Does equality of opportunity require inequality of liberty? These are at least in part answerable only by empirical inquiry and data from history and the social sciences. Consequently, theorists who address these sorts of questions often combine techniques from analytical philosophy (i.e. tools of logic, for analysing forms of language, and increasing argumentative clarity) with those from empirical social science.

Theorists might use any or all methods of gathering facts that empirical researchers use. A theorist addressing a question concerning existing beliefs might employ methods such as interviews, surveys, or the analysis of language or meaning in discourse and texts in order to chart expressed opinions and bring implicit assumptions to light. The research might engage with and incorporate the findings of sociologists, economists, and experts in social policy. It might seek to identify statistical correlations, and figure out causal mechanisms as well. The question might concern how a value—justice, efficiency, national solidarity,

welfare, security, democracy—ought to be understood. An argument might be developed concerning the implications of a particular conception of, or moral premise relating to, this value. It might be developed with reference to some particular policy or institutional area, and by appeal to logic and empirical evidence. A researcher might reflect on the relevance and applicability of a theory to the clarification and solution of a normative issue or problem in a particular domain. For example, the research might involve analysing the usefulness of a particular theory of justice, or comparing the utility of several theories, for the study of problems relating to globalization, migration, nationalism, multiculturalism, education, ageing, changing gender roles, or global inequality.

In general, researchers addressing normative and empirical questions will develop answers to them in broadly similar ways. As in any research, the normative political theorist finds and formulates an interesting question, reads what others have written on that question, engages with this literature and, through that engagement, formulates an argument of their own.

We have previously argued that, irrespective of what kind of question a researcher addresses (descriptive, explanatory, predictive, prescriptive, or normative), thinking in terms of formulating a hypothesis in answer to it can help to clarify the argument and the kinds of evidence that will provide a meaningful demonstration of it. A *normative hypothesis* advances, for further discussion, analysis and investigation, an argument or suggestive proposition about what ought to be. Recall that a hypothesis consists of three elements: an independent variable, a dependent variable, and a statement about the relationship between them. While we don't think of normative political theory as requiring a dependent and independent variable, normative questions nonetheless are concerned with connections and relations, with demonstrating relations between premises and conclusions, and indicating the nature of the relations between them. A normative hypothesis might state that the best type of X is Y; that the basis for deciding or justifying X ought to be Y, the legitimacy of X ought to be based on Y, or that X is just when conditions A, B, C are present. Investigation of these hypotheses would entail elaborating an argument about why this should be the case and establishing the plausibility and coherence of the argument, with the help, if relevant, of empirical or historical evidence.

We began this discussion by observing that normative political researchers are often silent about the methods they use. It is not a question of *whether* political theorists employ methods: they *do*. A variety of methods is used to address normative questions. What is at issue is their tendency not to reflect on the methods they use. Irrespective of what type of research is pursued, research should be systematic, self-aware, clear, and transparent. This enables others to evaluate the argument with respect to the kind and nature of its premises, the strength of the inferential links between premises and conclusions, and its possible criticisms or refutations.

Conclusions

We have organized our discussion of the various considerations and tasks involved in developing an answer to a research question around three basic requirements. The first requirement is that the answer be appropriate to the type of question that is being asked. The second requirement of an answer to a research question is that it makes a contribution to

knowledge. The third is that the answer must be clearly and fully specified with regard to the factors or elements you think must be taken into consideration in order to answer the question, and how you think these factors or elements are related to each other.

We maintain that for all types of research, answers or arguments in response to research questions benefit from being formulated as a hypothesis and an investigation of it. We determine which ideas are most useful for understanding social life by turning our ideas into hypotheses and assessing them. A hypothesis articulates the contours and logic of the investigation. In all research, a hypothesis helps to make explicit the argument that is being developed and to guide research. Hypotheses can be investigated either by testing them with evidence (confirmatory research) or through their operation as a guide to a process of discovery (exploratory research).

Next up: How do you know if your hunch is right? What kind(s) of evidence would give you confidence that your proposition is probably correct or, alternatively, lead you to conclude that the proposition is probably wrong?

Questions

1. What is the difference between a concept and a variable? How are the two related?
2. What is a theory?
3. What is the role of theory in political and social scientific inquiry? How does theory contribute to a furthering of our understanding of the political and social world?
4. In what ways is the study of concepts important for political and social scientific inquiry?
5. What is the function of operationalizing concepts?

Guide to Further Reading

Bell, D. and A. De-Shalit (eds) (2003), *Forms of Justice: Critical Perspectives on David Miller's Political Philosophy* (Lanham, MD: Rowman and Littlefield).

Political philosophers have traditionally argued that there is a single, principled answer to the question of what justice is. But David Miller (*Principles of Social Justice*, Cambridge, MA: Harvard University Press, 1999) has theorized that justice can take many different forms. Taking this position as a starting point, this collection of articles considers whether justice takes one form or many by drawing real-world implications from theories of justice and examining in-depth social justice, national justice, and global justice.

Collier, D. and H. Brady (eds) (2004), *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Lanham, MD: Rowman and Littlefield).

This is a set of essays that respond to King, Keohane, and Verba's *Designing Social Inquiry* (discussed in Chapter 1). The essays endorse the argument in King et al. concerning the need for common standards; but they place greater emphasis on the contribution of qualitative methodology and the need for diverse tools for inquiry. The volume provides an overview of *Designing Social Inquiry* and a chapter that draws together basic issues in the ongoing methodological debate.

Gerring, J. (2001), *Social Science Methodology: A Criterial Framework* (Cambridge: Cambridge University Press), 'Concepts: General Criteria' and 'Strategies of Definition', 35–86.

In this influential discussion of, and answer to, the question of what distinguishes a good concept from one that is less good or less useful, Gerring identifies and discusses eight essential criteria for a good concept.

Goertz, G. (2005), *Social Science Concepts: A User's Guide* (Princeton, NJ: Princeton University Press).
This provides an analysis of concepts and their role in the research enterprise, as well as detailed advice on how to construct them.

Leopold, D. and M. Stears (eds) (2008), *Political Theory: Methods and Approaches* (Oxford: Oxford University Press).
This explores the ongoing debate among political theorists concerning the relationship between normative political theory and empirical inquiry, and introduces key approaches to political theory used by leading scholars in the field.

Sartori, G. (ed.) (1984), *Social Science Concepts: A Systematic Analysis* (New York: Sage).
Sartori's theoretical chapter, entitled 'Guidelines for Concept Analysis', explains how words acquire multiple meanings and provides a succinct discussion of the problems involved in defining concepts and finding suitable terms for them. It contains chapters examining key terms, for example consensus, development, ethnicity, integration, political culture, power, and revolution. In each of these chapters, the starting point is a word that has acquired a multiplicity of meanings.

Schmitter, P. (2008), 'The Design of Social and Political Research', in D. Della Porta and M. Keating (eds), *Approaches and Methodologies in the Social Sciences: A Pluralist Perspective* (Cambridge: Cambridge University Press), 263–95.
This is an excellent discussion of research design—its relationship to the research question and other components of the research process, and the considerations that must be taken into account in constructing one.

Schramme, T. (2008), 'On the Relationship between Political Philosophy and Empirical Sciences', *Analyse und Kritik* 30: 613–26.
Schramme argues that the findings of the empirical sciences might play a role in justifying normative claims in political philosophy. He describes how political theory has become a discipline divorced from empirical sciences, and outlines some functions that empirical studies might have in political philosophy.

References

Alvarez, M., J. A. Cheibub, F. Limongi, and P. Przeworski (1996), 'Classifying Political Regimes', *Studies in Comparative International Development* 31: 3–36.

Audi, R. (1983), 'The Applications of Conceptual Analysis', *Metaphilosophy* 14(2) (April): 87–106.

Bachrach, P. and M. S. Baratz (1962), 'Two Faces of Power', *American Political Science Review* 56(4) (December): 947–52.

Bollen, K. A. (1980), 'Issues in the Comparative Measurement of Political Democracy', *American Sociological Review* 45 (June): 370–90.

Bollen, K. A. (1993), 'Liberal Democracy: Validity and Method Factors in Cross-National Measures', *American Journal of Political Science* 37: 1207–30.

Bollen, K. A. and R. W. Jackman (1989), 'Democracy, Stability, and Dichotomies', *American Sociological Review* 54: 612–21.

Bollen, K. A. and P. Paxton (2000), 'Subjective Measures of Liberal Democracy', *Comparative Political Studies* 33(1) (February): 58–86.

Carens, J. (2000), *Culture, Citizenship, and Community* (Oxford: Oxford University Press).

Cohen, G. A. (2008), *Rescuing Justice and Equality* (Cambridge, MA: Harvard University Press).

Collier, D. and S. Levitsky (1997), 'Democracy with Adjectives: Conceptual Innovation in Comparative Research', *World Politics* 49(3) (April): 430–51.

Collier, D. and J. E. Mahon (1993), 'Conceptual "Stretching" Revisited: Alternative Views of Categories in Comparative Analysis', *American Political Science Review* 87(4) (December): 845–55.

Coppedge, M. and W. H. Reinicke (1990), 'Measuring Polyarchy', *Studies in Comparative International Development* 25: 51–72.

Corbin, J. and A. Strauss (1990), 'Grounded Theory Research: Procedures, Canons and Evaluative Criteria', *Qualitative Sociology* 13: 3–21.

Dahl, R. A. (1957), 'The Concept of Power', *Behavioral Science* 2: 201–15.

Dahl, R. A. (1971), *Polyarchy: Participation and Opposition* (New Haven, CT: Yale University Press).

Dahl, R. A. (1989), *Democracy and its Critics* (New Haven, CT: Yale University Press).
Dunn, J. (1990), 'Reconceiving the Content and Character of Modern Political Community', in J. Dunn (ed.), *Interpreting Political Responsibility* (Oxford: Polity), 193–215.
Farrelly, C. (2007), 'Justice in Ideal Theory: A Refutation', *Political Studies* 55 (December): 844–64.
Gerring, J. (1999), 'What Makes a Concept Good? A Criterial Framework for Understanding Concept Formation in the Social Sciences', *Polity* 31(3) (spring): 357–93.
Gerring, J. (2001), *Social Science Methodology: A Criterial Framework* (Cambridge: Cambridge University Press).
Glaser, B. G. and A. L. Strauss (1967), *The Discovery of Grounded Theory* (Chicago: Aldine).
Gurley, J. G. (1975), *Challengers to Capitalism: Marx, Lenin, and Mao* (Stanford, CA: Stanford Alumni Association).
Guzzini, S. (1993), 'Structural Power: The Limits of Neorealist Power Analysis', *International Organization* 47(3): 443–78.
Hoover, K. and T. Donovan (2004), *The Elements of Social Scientific Thinking*, 8th edition (Belmont, CA: Wadsworth/Thompson).
Jacobson, G. (1978), 'The Effects of Campaign Spending in Congressional Elections', *American Political Science Review* 72 (June): 469–91.
Kahl, C. H. (2006), *States, Scarcity, and Civil Strife in the Developing World* (Princeton, NJ: Princeton University Press).
Kumar, R. (2005), *Research Methodology: A Step-by-Step Guide for Beginners*, 2nd revised edition (London: Sage Publications Ltd).
Lenin, V. I. (1916), *Imperialism, the Highest Stage of Capitalism: An Outline* (New York: International Publishers [1939]).
Leopold, D. and M. Stears (2008), 'Introduction', in D. Leopold and M. Stears (eds), *Political Theory: Methods and Approaches* (Oxford: Oxford University Press), 1–10.
Levy, J. S. (1990–1), 'Preferences, Constraints, and Choices in July 1914', *International Security* 15(3) (Winter): 151–86.
Lukes, S. (1974), *Power: A Radical View* (New York: Palgrave Macmillan).
Mason, A. (2004), 'Just Constraints', *British Journal of Political Science* 34(2): 251–68.
McDermott, R. (2001), *Risk-Taking in International Politics: Prospect Theory in American Foreign Policy* (Ann Arbor: University of Michigan Press).
Merton, R. K. (1968), *Social Theory and Social Structure* (Glencoe, IL: Free Press).
Miller, D. (2008), 'Political Philosophy for Earthlings', in D. Leopold and M. Stears (eds), *Political Theory: Methods and Approaches* (Oxford: Oxford University Press), 29–48.
Nye, J. S. (1990), *Bound to Lead: The Changing Nature of American Power* (New York: Basic Books).
Nye, J. S. (2002), *The Paradox of American Power: Why the World's Only Superpower Can't go it Alone* (New York: Oxford University Press).
Nye, J. S. (2004), *Soft Power: The Means to Success in World Politics* (New York: Public Affairs).
Pollins, B. (2007), 'Beyond Logical Positivism: Reframing King, Keohane, and Verba', in R. N. Lelbow and M. I. Lichbach (eds), *Theory and Evidence in Comparative Politics and International Relations* (Basingstoke: Palgrave/Macmillan), 87–106.
Rawls, J. (1971), *A Theory of Justice* (Cambridge, MA: The Belknap Press of Harvard University Press).
Sartori, G. (1970), 'Concept Misformation in Comparative Politics', *American Political Science Review* 64: 1033–53.
Sartori, G. (1984), 'Guidelines for Concept Analysis', in G. Sartori (ed.), *Social Science Concepts: A Systematic Analysis* (Beverly Hills, CA: Sage), 15–85.
Sartori, G. (1987), *The Theory of Democracy Revisited* (Chatham, NJ: Chatham House).
Schmitter, P. (2008), 'The Design of Social and Political Research', in D. Della Porta and M. Keating, *Approaches and Methodologies in the Social Sciences: A Pluralist Perspective* (Cambridge: Cambridge University Press), 263–95.
Sen, A. (2009), *The Idea of Justice* (Cambridge, MA: Harvard University Press).
Shively, W. P. (1989), *The Craft of Political Research*, 2nd edition (Englewood Cliffs, NJ: Prentice-Hall, Inc.).
Stears, M. (2005), 'The Vocation of Political Theory: Principles, Empirical Inquiry and the Politics of Opportunity', *European Journal of Political Theory* 4(4): 325–50.
Strange, S. (1994), *States and Markets*, 2nd edition (London: Pinter).

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

6 Research Design

Chapter Summary

Previous chapters have focused on how to find and formulate a research question (Chapter 4) and how to develop a hunch about the answer to one (Chapter 5). In this chapter, we focus on designing research to investigate your hunch. Our discussion focuses on the following issues:

- basic principles of research design;
- types of research design;
- selecting cases and units of observation and analysis;
- types of evidence researchers use to answer the questions they ask;
- methods of collecting data;
- sources of data.

Introduction

You have a well-crafted research question. It is one that you have shown (1) has significance for a topic or issue relating to the subject matter of our field; (2) is researchable (it can be answered through conducting research); and (3) has not yet been answered definitively. You also have a hunch about the factors you think must be taken into consideration in order to answer your question, and how you think these factors are related to each other. Together, these factors and their relationships constitute what we call a '*hypothesis*': a reasoned, clearly specified hunch or expectation about the answer to your question that will guide and focus your research. In Chapter 5 we said that your response to a research question will consist of a hypothesis and an investigation of it; and we discussed the basic components of a hypothesis, and how to formulate one. We turn now to a discussion of how to investigate it.

So you have a hunch or argument about the answer to your research question. How do you know whether your hunch is right? And how will you convince others that it is right? Your strategy for providing a convincing 'test' or demonstration of your hypothesis is what we call a 'research design'. A research design specifies the sort of evidence you need to investigate your hypothesis, and describes how the evidence will be collected and analysed.

In this chapter we first discuss the basic principles of research design. We then provide an overview of (1) types of research design; and (2) methods of data collection. Research designs and data-collection methods are often confused with each other, and indeed there is a certain amount of overlap between them. The overall strategy you will employ to

demonstrate or investigate your argument is your *research design*. The sort of data that will enable you to implement this strategy, and where and how they will be obtained, is your *method of data collection*. A research design answers the question, 'What sort of evidence will test my hypothesis?' The answer might be 'data on the perceptions, attitudes, and opinions of politicians'. A data-collection method answers the question, 'How and where can I get this information?' It might specify, for instance, that we will use surveys or interviews, ethnographic research, or content analysis, or all four together. It will specify whether the data will be collected using qualitative or quantitative methods, and whether we will collect them ourselves or use data collected by others.

In some fields, the term 'research design' is also used to refer to the overall plan for addressing a research problem or question. In this usage, the 'research design' includes the definition of the research question, the hypotheses to be examined, including the number and type of variables to be studied and the relationships between them, and the schedule and budget for carrying out the research. While a research design and a data-collection method are closely related in the sense that a particular method of data collection must fit with the strategy defined by a research design, they represent two separate steps or components in the research process.

Often students will come to us and say something along the lines of the following: 'I am going to be in Egypt over the semester break and, as I want to write my thesis on Egyptian foreign policy, I was wondering whether you think it would be a good idea to try and conduct some interviews while I am there?' In most cases our response will be, 'I don't know:

BOX 6.1 Components of the Research Process

Research components	What the component involves
The research question	Formulating a research question and providing a rationale for its selection.
↓	
The literature review	Critically discussing and synthesizing previously published literature associated with the question.
↓	
The theoretical framework	Elaborating your own answer/argument/hypothesis in response to the question.
↓	
The research design	**Constructing a design or blueprint for carrying out research—a strategy for investigating the hypothesis in a coherent and logical way. The strategy will involve identifying the sort of data that will be necessary for an adequate testing of the argument or hypothesis and serve as a basis for drawing inferences; how such data will be obtained; and what methods of analysis will be applied to the data to determine whether or not the hypothesis is true or false.**
↓	
The actual research	Carrying out the process of data collection and analysis that the research design has mapped out.

it depends on the research question you intend to address and what hypothesis you decide to investigate.'[1]

The point is, you cannot collect data or gather information until you know precisely what sort of evidence you need. What you will need to be looking for is evidence that would confirm or disconfirm the argument you plan to advance or investigate; but you won't know what sort of evidence that *is*, unless you know what argument or hypothesis you are putting forward for investigation. To call upon the familiar 'toolbox' metaphor, it is as if you said 'There is something in the basement that we want to repair. Should we use a hammer or a wrench?' How can we know which tool to suggest if we don't know the nature of the problem and how you think you might go about solving it?[2] *First*, you formulate a research question and a hunch about what might provide an answer to it; *then* you can decide what information you will need to collect in order to confirm or disconfirm your hypothesis. Only then will you know whether it would be useful to conduct interviews, and, if so, who you should interview and what you should ask.

In Chapter 1 we said that the research process—the process of asking and answering questions about the political world—involves a number of steps. Box 6.1 shows where 'research design' fits into this process.

Basic principles of research design

Before you begin your research, you will need to think carefully and creatively about what sort of research will enable you to answer your question or test your hypothesis *in a convincing way*. Your research design is the culmination of this thinking. It specifies what will enable you to draw logical, valid, and reliable conclusions about your hypothesis. A good research design has the following attributes:

1. It specifies the type of research and techniques of data collection appropriate to the objectives of the project.
2. It makes explicit the logic which enables you to draw inferences—logical conclusions based on the information you collect or observations you make.
3. It identifies the type of evidence that not only confirms your hypothesis or argument, but provides a convincing 'test' of it.
4. It decreases threats to the internal and external validity of your findings.
5. It ensures that your findings are reliable.

Let's clarify what each of these principles requires you to consider.

Principle 1: A research design *specifies the type of research and techniques of data collection appropriate to the objectives of the project*.

Different types of research design and methods of data collection are not more or less rigorous than others; they are only more or less appropriate for the task at hand. A saw is not better or worse than a wrench; it is only more or less useful in relation to a specific task. A research design is good if it allows researchers to draw valid inferences. What that means is clarified by the the second principle, below.

Principle 2: A research design *makes explicit the logic which enables you to draw inferences—logical conclusions based on the information you collect or observations you make.*

A research design should provide a structure of inquiry that enables us to draw *logical* conclusions on the basis of known facts or premises; and it should make explicit the logic linking the facts we know (that we collect or observe) to those we don't (those that are the subjects of our research questions and hypotheses). The process of using the facts we know to draw logical conclusions about facts we do not know is what we call **inference** (King, Keohane, and Verba 1994: 46). This is the basic logic that underlies most thinking. However, in conducting research, the process must be made explicit and must follow certain rules.

Your research must be presented in a way that is clear, orderly, and systematic enough that someone else will be able to retrace your steps and arrive at the same results and conclusions. Knowledge is the product of research that is open at every step to checking and double-checking by others. Consequently, you must be both transparent and explicit. You must describe what you did (or plan to do) in clear, simple language.

Principle 3: A research design *identifies the type of evidence that not only confirms your hypothesis or argument, but provides a convincing 'test' of it.*

The components of a research project should be designed to provide a convincing test of your hypothesis. As David de Vaus puts it, rather than seeking *evidence that is consistent* with our argument or hypothesis, we should seek *evidence that provides a compelling test* of it (de Vaus 2001: 11). You can usually find empirical support for almost any hypothesis. So rather than marshalling evidence to fit an argument or hypothesis, you should seek evidence that can 'test' it against rival perspectives, arguments, or hypotheses (see the discussion of 'literature reviews' in Chapter 4) and show why your explanation or interpretation is to be preferred. Silence your sceptics by raising the objections or questions that they might raise. Assume their point of view and argue, on their behalf, against yourself. Think carefully about what sort of evidence will enable you to make comparisons and judgements between alternative possible explanations, and to provide the analytic leverage you need to demonstrate that your hypothesis is in some way better, more credible, useful, accurate, or illuminating than its rivals.

Principle 4: A research design *decreases threats to the internal and external validity of your findings.*

The better the research design, the more it decreases threats to validity. A study is valid if its measures actually measure what they claim to measure, and if there are no logical errors in the conclusions drawn from the data. Researchers distinguish among a variety of different types of validity. We will discuss two types: internal validity and external validity.

Internal validity often refers to the extent to which we can be confident that the independent (causal) variable produced the observed effect. The more the structure of a study eliminates alternative interpretations, the better we are able to draw unambiguous conclusions from our results, and 'the stronger the internal validity of the study' (de Vaus 2001: 28). What is at issue is whether there are factors, other than your independent variables, that might be affecting the outcome. Is the independent variable responsible for variation in the dependent variable? What other possible causes might there be for the relationship between the variables? Could something else have been responsible for the variation in the dependent variable? Could there be confounding factors? These are the biggest threat to validity: if you do not design your research in a way that enables you to rule out other factors

or alternative explanations, the internal validity of your study will be threatened. **External validity** refers to 'the extent to which results from a study can be generalized beyond the particular study' (de Vaus 2001: 28). Can you generalize your findings? Are your conclusions likely to apply more widely? Are they applicable to other similar situations or cases?

Principle 5: A research design *ensures that your findings are reliable*.

Research should be designed to ensure that your findings are reliable. **Reliability** refers to the 'repeatability' or 'consistency' of your findings. A research design is reliable if other researchers can perform exactly the same procedures and come up with the same results (your findings are repeatable). A measure is reliable if it gives us the same result over and over again (assuming that what we are measuring isn't changing).

Now that we know what a research design does for us, we can turn to a consideration of different types of research design used in political research. Before moving on, however, you may want to take a look, in Box 6.2, at some of the terms that this chapter has introduced so far.

BOX 6.2 Key Terms

Research Design	A logical structure of inquiry that specifies the sort of test or evidence that will convincingly confirm or disconfirm a hypothesis.
Data-Collection Methods	Strategies for gathering information.
Inference	The process of drawing logical conclusions on the basis of premises assumed to be true.
Internal Validity	The extent to which we are confident that a causal relationship really exists between two variables.
External Validity	The extent to which findings of your research can be generalized to other similar situations.
Reliability	The extent to which a measurement produces the same results on repeated trials.

Types of research design

The types of research design that are commonly used in our field include experimental designs, cross-sectional and longitudinal designs, and comparative designs. We first discuss experimental designs, which we will discuss further in Chapter 8. Following this, we outline the basic principles of cross-sectional and longitudinal designs. We then provide a brief introduction to comparative designs, which we take up in greater detail in Chapter 9, and historical research, which we explore more fully in Chapter 10.

Experimental designs

Experimental designs are often thought to most closely resemble the true scientific method (see Chapter 2), and as such they are widely regarded as being the most effective design for testing whether or not two variables are causally related. They manage to do this thanks to

the rigorous use of experimental control. This helps to overcome one of the main problems that researchers face when they want to investigate causal relationships: there are a vast number of potentially important variables that may influence any given political phenomenon, and in the absence of rigorous controlled experiments (not dissimilar to those used in medical research) it is very difficult to know which are responsible for causing the phenomenon under investigation. One of the great strengths of experimental research is that the researcher can control the environment and manipulate particular variables of causal interest with great precision. Indeed, the defining characteristic of experimental research is intervention by the researcher in the data-gathering process (Morton and Williams 2010). We call the data gathered by intervention (or treatment, as it is sometimes known) experimental data.

Broadly speaking there are three main ways of carrying out experimental research: in a laboratory, in the field, and by utilizing natural occurrences. In *laboratory experiments*, the **subjects** are recruited to a common location where the experiment takes place. The laboratory experiment is designed to ensure that the researcher has as much control as possible over the environment to which subjects are exposed. In this way the experimental group and the control group can be exposed to exactly the same environment except for the experimental intervention. Accordingly, any difference that is recorded on the post-test measurement of the dependent variable can be confidently attributed to the presence (or absence) of the intervention. However, because of concerns about the artificial environment of laboratory experiments, some political scientists favour the more natural setting of field experiments. In *field experiments*, the intervention by the researcher takes place in real-world environments, and subjects may not even know that they are participating in an experiment. This obviously raises some ethical issues to do with informed consent and deceit (which we discuss in detail later in this chapter). But as Green and Gerber (2003: 94) observe, field experiments have two main strengths. First, random assignment ensures unbiased inference about cause and effect, and second, the natural settings ensure that the results will tell us something useful about the real world.

The third main type of experimental design is often referred to as a natural experiment. *Natural experiments* rely on naturally occurring events as interventions rather than interventions controlled by the researcher. Even though in this case the researcher is not doing the intervening, the approach taken with the data is as if the researcher has intervened.

There is often thought to be a trade-off between these different types of experimental designs. In particular, it is often argued that laboratory experiments tend to have relatively high levels of internal validity but low levels of external validity, whereas field experiments tend to have higher levels of external validity but lower levels of internal validity. That is, with lab experiments, the controlled setting in which the experiment takes place enables the researcher to have a high degree of confidence in the effect of the causal variable on the outcome of interest (high internal validity). However, at the same time, because of the artificial environment of the laboratory setting, it is less certain whether the findings from the experiment may be applied or generalized to real-world settings (low external validity). By contrast, since field experiments take place in the real world and reproduce as closely as possible the conditions under which different political phenomena occur, they tend to have higher levels of external validity, meaning the findings can be generalized with more confidence.

Given these different strengths and weaknesses, rather than prioritizing one approach over the other, or regarding one approach as inherently superior to the other, it makes more sense to think about how the two approaches can be combined. For a study to really stand up, it should be possible to try and examine it in both settings. As Elinor Ostrom (2007: 26–7) says, 'To test theory adequately, we need to use methods that together combine external and internal validity. One gains external validity in doing field research, but internal validity in the laboratory.'

Although experimental research has certain obvious strengths, particularly when it comes to testing causal hypotheses, it is still not used particularly widely in political research. However, there is the potential for it to be used much more widely than it is. In Chapter 8 we discuss some of the main reasons why it is not used more widely, with reference to some of the restrictions that ethical issues and practical considerations play in determining what sort of questions can be answered. But we also go on to discuss the potential avenues for future research.

However, even if experimental research is not used widely, it has still been incredibly influential in terms of shaping how we can study politics. Most of our research designs are, in effect, an effort to approximate the logic of experimental design as closely as possible, and in the positivist hierarchy large-N quantitative analysis is often seen as the next best thing to doing experimental research.

Cross-sectional and longitudinal designs

Cross-sectional designs involve analysing a sample, or cross-section, of a population at a single point in time. **Longitudinal designs** explore changes or trends over time. In a longitudinal study, research is repeated with the same sample or set of cases over two or more intervals. To illustrate how these two designs might differ in how they address the same question, consider the hypothesis that as countries get richer income distribution becomes more unequal. A researcher might employ either a cross-sectional research design or a longitudinal design to investigate this hypothesis. Using a cross-sectional design, a researcher might select a sample of countries with different levels of economic growth (low, medium, high) and analyse the association between level of growth and change in inequality. With a longitudinal design, a researcher might select a country or countries that have moved from lower to higher levels of development to see whether income distribution *over time* was consistent with the trend suggested by the hypothesis.

A key feature of a cross-sectional design is that it is concerned with explaining variation between places or between people at a single point in time, rather than explaining variation within places over time. An example of a cross-sectional study is one conducted by Timmons Roberts, Bradley Parks, and Alexis Vasquez to investigate 'Who Ratifies Environmental Treaties and Why?' (2004). In this study, Timmons Roberts and his co-authors try to explain variation between different states' behaviour with respect to international environmental issues. To explore this question they conducted a cross-sectional study on the determinants of whether or not states have ratified international environmental treaties, using data from 177 countries. They found that most of the variance in the ratification of environmental treaties is accounted for by three variables: 'disadvantaged insertion into the world economy' (defined in terms of countries with a narrow export base), voice

and accountability through domestic institutions (the degree to which citizens choose those who govern them and the independent role that the media plays in keeping government accountable), and civil society pressure (number of non-governmental organizations (NGOs) in the country). In particular, 'the number of NGOs in a nation appears virtually synonymous with its likelihood to participate in environmental treaties' (2004: 39). While the dependent variable for this study is a measure of how many environmental treaties were ratified between 1946 and 1999, the independent variables (a country's position in the world economy, voice and accountability, number of NGOs per country) are drawn from a single point in time (mostly from 2000).

When data are collected at *a single point in time*, as in a cross-sectional study, analysts must be careful about drawing any conclusions about *changes over time*. In order to overcome this problem, repeated cross-sectional studies can be used to introduce a longitudinal element into the research design. An example of a repeated cross-sectional study is William Mishler and Richard Rose's (2007) investigation into how citizens, socialized by authoritarian regimes in quintessentially authoritarian cultures, can learn the attitudes and behaviours necessary to become loyal and effective citizens of new democratic regimes. Mishler and Rose address this question by constructing a repeated cross-sectional study to test two competing hypotheses drawn from two theories that dominate the study of processes of political learning and relearning.

Cultural theories hypothesize that basic political attitudes inculcated through early life socialization are deeply ingrained and change only slowly over extended periods. *Institutional theories* emphasize adult political experiences or political 'relearning' based on individuals' rational assessments of contemporary institutions and circumstances. Attitudes and behaviours are malleable and adaptable. 'Thus, later life experiences are expected to play a greater role in shaping adult opinions' (Mishler and Rose 2007: 823). The two theories yield different predictions: cultural theories 'predict substantial generational differences in political attitudes … that change little over an individual's lifetime'; institutional theories 'predict substantial changes among individuals across either generations or the life cycle' (Mishler and Rose 2007: 824). The authors test these competing perspectives using data from a series of national probability surveys conducted in Russia between 1992 and 2005. Each survey is an independent cross-section of the population and not part of a panel design (one in which a particular set of respondents are questioned repeatedly). Although the content of the surveys changes over time to reflect changing conditions in Russia, they include a core of questions asked consistently over time to facilitate comparisons. Their evidence suggested that 'individual assessments of contemporary political and economic experiences' have much greater effects on regime support than do generational effects, and that 'Russians' lifelong socialization into an authoritarian culture by an authoritarian regime is not in itself an insurmountable obstacle to the development of democracy in Russia' (Mishler and Rose 2007: 832).

While repeated cross-sectional studies can be used to measure trends or 'aggregate' change over time, they don't measure 'individual' development or change. This type of design might therefore be less satisfactory for investigating, for example, the relationship between individuals' civic orientations and political participation. As Jan Leighly points out, studies that rely on cross-sectional survey data on individuals' characteristics, such as civic orientations, as predictors of participation assume that positive civic orientations are

causally prior to acts of participation. But the direction of causality can run in the opposite direction: participation can lead to changes in individuals' political attitudes, and particularly in their sense of political efficacy and sophistication (1995: 186). Studies of political participation therefore need to look at individuals' acts of participation, and track how their participation decisions change over time (1995: 198). This is what longitudinal designs allow us to do. They study processes of individual development and change over time, and the effects of earlier events or characteristics on later outcomes.

There are two main types of longitudinal design. In a 'cohort' study, the researcher selects a group of people within a delineated population—that is, having a similar characteristic or experience (e.g. an age or ethnic group)—and charts the individuals' development processes from a particular time point. In a panel study, the researcher chooses a sample, often a randomly selected national one, and collects information on it at two or more points over time. While panel studies gather information on the same people at each time point, the number of participants tends not to be constant over time, because of attrition. This can introduce bias into the results. Longitudinal data are particularly useful in answering questions about the dynamics of change. For example, under what conditions do voters change political party affiliation? What are the respective roles of mass media and friends in changing political attitudes? It is also useful in predicting long-term or cumulative effects that are normally hard to analyse in a cross-sectional study.

Comparative designs

Comparative research designs are perhaps the most widely used research design in political research. In a sense everything is comparative. We can have comparative experimental designs, comparative cross-sectional designs, and comparative longitudinal designs. We can even think of the single country case study as a type of comparative design, since it usually involves a comparison of some type or another, whether it is between regions within the country or between periods over time. Within the comparative framework it is common to distinguish between three main types of research design. There are large-N studies (where N refers to the number of countries—or cases—that are compared), small-N studies (involving the analysis of a small number of countries, typically 2, 3, 4 … but with no real upper limit), and single-N studies (otherwise known as case studies). These designs are distinguished primarily in terms of how many countries (or cases) are compared, but also in terms of how the countries for analysis are selected. Both aspects of case selection are very important. These are discussed in detail in Chapter 9.

There is often thought to be a trade-off between the in-depth, intensive knowledge derived from the study of a small number of cases, on the one hand, and the extensive, cross-case knowledge based on the study of a large number of cases, on the other—although this is often overstated. Small-N studies or single-country case studies are often used to uncover causal paths and mechanisms and assess specific mechanisms identified in theories. This is frequently referred to as **process tracing** (more on this, in Chapter 10). Research using a large number of cases may observe a strong statistical relationship between two variables, and may use a theory to describe these statistical results. But if the researcher is unable to observe directly the key mechanisms of the theory, it will be difficult to know if the mechanisms producing the statistical relationship are the same as those described in the theory.

Selecting a small number of cases for in-depth investigation can enable a researcher to test for the existence of these mechanisms. The study of a small number of cases can also enable a researcher to investigate a case that is theoretically anomalous. Detailed study of a case that deviates from theoretical expectations may generate findings that lead us to substantially revise or altogether discard existing theories. Explaining anomalies can also be a source of new theories. In-depth investigation of a small number of cases can also help in generating hypotheses and theories in developing fields of inquiry.

Because case study designs can incorporate a broader history and wider context than can other designs, case study research is generally strong in dealing with two threats to internal validity: the threats which 'history' and 'maturation' present to the validity of a study. 'History' refers to historical or contextual factors, and 'maturation' to natural changes that affect the relationship between the independent variables and the outcome. Case study research is generally weaker in external validity (i.e. generalizability) because it includes only a small number of cases of some more general phenomenon (Gerring 2007: 43). Greater generality can be achieved by using a larger set of cases (as, for instance, in comparative case study designs). However, extending the analysis to broader contexts might lead to conceptual stretching, and thus threaten the conceptual validity of the study (Collier and Mahoney 1996: 69; see Chapter 5 for a discussion of 'conceptual stretching').

Research costs will also be a factor in the number of cases you select to study. If each case requires a limited amount of easily collected information, you might include many, or even all, relevant cases. If you need a great deal of information, or the information you need is harder to collect, you will examine fewer cases—perhaps only one or two. Cross-national studies of war that rely on quantitative measures that are easily computed from available statistical sources (e.g. United Nations annuals or computerized databanks) might include every country in the world. But a study of the impact of privatization policies on social inequalities among multi-ethnic populations might require more detailed information that can be gathered only through close analysis, and would lead the researcher to focus on only a few cases.

You must provide a rationale for why the specific case or set of cases you selected, from among all those in the larger population, were chosen. Researchers may select cases because they are *critical* (to testing a theory), *revelatory* (reveal relationships which cannot be studied by other means), or *unusual* (throw light on extreme cases) (Yin 1984).

But whatever the purpose they are chosen to serve, it is important that they are selected with care, and that the selection is based on the type of case that will provide the most convincing test or investigation of your hypothesis. Ultimately, the choice of how many and which cases you study and how you will study them will be determined by your research question and the hypothesis that you intend to investigate, though it is becoming more common to integrate these different research designs together, so as not to leave out either the general or the specific.

Historical designs

Historical research designs can help us to discover the influence of certain events on others. They do this in two ways: first, by contextualizing events so as to enable us to better understand the relationships between those events; and, second, by helping us to understand how

the timing and sequence of social actions and processes impact on their nature and on the development of subsequent actions and events.

Researchers in our field employ a number of different historical designs for this purpose. The first of these, *historical events research*, involves the analysis of one case at one period in time. This type of research often focuses on an event that has had an important impact on subsequent events or has important implications for a general theory. For instance, a landmark study by Graham Allison (1969) focused on a single historical case, the Cuban Missile Crisis of 1962, to test the theoretical assumptions of the rational actor model of foreign policy decision-making and to develop models based on alternative assumptions. *Cross-sectional comparative research* involves the study of many cases from one time period, as for instance Paul Pierson's (1994) influential study comparing the dismantling of the welfare state during the Thatcher/Reagan period. This type of research design is similar to the comparative designs discussed previously.

Historical research designs can also be longitudinal. *Historical process research* focuses on a series of events within a *single* case in order to understand how processes change over time. *Comparative historical research* focuses on a series of events across *many* cases in order to understand how processes change over time and between places. This has become a prominent research tradition in our field and, as Chapter 10 shows, has been used to study a wide range of issues.

Other historical research designs include *historical institutionalism* and *event structure analysis*. Historical institutionalism helps to explain how and why institutions originate, persist, and evolve. Researchers have used historical institutionalism to explore elements of persistence and change in a wide variety of institutions, including the modern state, political parties, organized labour, international organizations, and international law, to name but a few. Historical institutionalism is closely related to **comparative historical analysis**. But historical institutionalism is concerned more specifically with bringing questions of timing and temporality to the centre of the analysis of *institutions*, something which is the concern of many, but not all, comparative historical analyses. And while there are many historical institutionalist studies that engage in systematic comparison, many do not.

Historical institutionalism has developed a number of distinctive conceptual tools for exploring how timing and temporality shape political outcomes. Among the most important of these concepts is that of 'critical junctures' and the associated concepts of 'path dependence' and 'positive returns' or feedback. Process tracing is often employed in historical institutionalist studies because it can be used to uncover causal paths and mechanisms by tracing the sequence of events constituting a process (e.g. institutional persistence or change). *Event structure analysis* takes this further by helping to make explicit the causal logic of sequences. It does this through an analytic procedure that enables researchers to probe the links between actions that constitute the chronology of a given event for potential causal ties among those actions. These are discussed in detail in Chapter 10.

Knowing about the context within which political institutions, structures, and processes emerged, and the founding moments, or 'critical junctures', which set in motion a process or direction of development helps us to address all types of questions. It helps researchers to better understand the characteristics of something, or to describe how it works or behaves. Understanding the historical context of something that has occurred or is happening also enables researchers to explain what factors or conditions are causally connected to it.

Historical research also provides analytic leverage in addressing predictive and prescriptive questions. Predictive questions ask about the likely outcome of existing trends. Will ethnic conflict increase in European cities in the next twenty years? What impact will structural changes in energy systems have on the Middle East in the next twenty years?

Historical research helps researchers to address these questions because it enables them to draw comparative insights into contemporary issues by exploring cases for which we have a fuller, more complete record and a larger body of scholarly work. So, for instance, a researcher interested in the question of whether China and Japan will develop more cooperative, or more conflictual, relations over the course of the next twenty years might explore the development of the German–French relationship following the Second World War, in which increasing economic and security interdependence led to an easing of historical tensions and greater cooperation between the two countries. They might then argue that, based on this history and the growing economic and security interdependence between China and Japan, it is reasonable to expect that China and Japan will probably develop more cooperative relations over the course of the next two decades, despite various issues which are currently causing tensions between the two countries.

Types of research, types of design

Empirical questions (descriptive, explanatory, or predictive questions) are concerned with how something works, why things happened the way they did, or what might be the outcome of current trends or conditions. An answer to a *descriptive question* might involve describing the characteristics of something; or modelling how it works or behaves. Common designs for these questions are case studies that provide an in-depth understanding of a process, event, or situation. *Explanatory questions* need a design that enables the researcher to determine whether one or more variables causes or affects one or more outcome variables, and to rule out other feasible explanations. *Predictive questions* require research designs that enable researchers to forecast future developments, based on an analysis of current events and relevant theories. We can analyse evidence and find that event A causes or explains event B (an *explanation*), and then predict that if A continues to increase, we are likely to have a greater amount of B in the future (White and Clark 1983: 23). Predictive questions require you to show that if certain conditions or circumstances prevail, a certain outcome is likely to occur or come into being.

Prescriptive questions are concerned with finding information to solve political problems. To investigate a hypothesis about what would be best to do in response to a problem, the researcher will inventory available options, weigh the pros and cons of each for achieving the desired outcome, and, based on this analysis, advance an argument for a policy using existing theoretical and empirical research (e.g. case studies). The researcher clearly defines the question to be answered or problem to be resolved; establishes relevant evaluation criteria so as to be able to compare, measure, and select among alternatives (e.g. cost, net benefit, effectiveness, efficiency, equity, administrative ease, legality, political acceptability); identifies alternative policies; and evaluates the costs and benefits of each, and the degree to which criteria are met in each of them.

Normative questions are usually settled through reflection on which view, when fully developed, offers the most coherent and convincing argument (Rawls 1993: 53). The 'test' of a

normative hypothesis consists of two interrelated components: (1) the logical consistency of the ideas—the conclusion and all points leading up to it must follow from the original premises; and (2) consideration of the arguments of other theorists who have written about these issues, both those that provide support for your own and those that raise criticisms and concerns that you will need to address if you are going to make a convincing case for your point of view. The appropriate research design, therefore, involves tracking down, as far as possible, persuasive arguments that certain conclusions follow logically from certain basic principles—principles over which, we can reasonably hope, far-reaching agreement can be reached. The method is to render such arguments—in respect to both basic and derivative principles—as clear, open, and logical as possible.[3]

In normative political analysis, the chief concern is with advancing an argument that is clear, logically sound, and rationally convincing. The researcher finds an interesting question, reads what others have written on the subject, engages with these works, and forms a carefully reasoned argument in reaction to them, with the help, if relevant, of empirical or historical evidence. Depending on the research question, the research might be analytic, critical, genealogical, deconstructive, or interpretive, and may draw on methods employed in analytical or applied philosophy, or the history of political thought. The aim might be to provide conceptual frameworks for deliberation, to analyse the logical structure of a principle, or to record and systematize the reasons that can validly be advanced for and against particular choices. In these cases, the researcher might employ the philosophical method of conceptual analysis, logic, and classification. Or the aim might be to draw out the implications of a moral premise within some limited problem area, appealing to logic and empirical evidence in varying degrees in the process. Some researchers might be concerned with the interpretation of existing beliefs. Their aim is to ascertain what people actually think by charting expressed opinions and bringing implicit assumptions to light. The analysis might be concerned to uncover and examine a value—justice, efficiency, national solidarity, welfare, security, democracy—that informs political policies and institutions, and offer an interpretation of how that value ought to be understood. In these cases, the research might employ interpretive methods concerned with working out exactly what a text is saying and why it's saying it in the way it does.

Data-gathering strategies: how the data are collected

Part of your research design involves determining what data and which data sources you will use to answer your research question. You should describe the data required to investigate your hypotheses and explain how and where you will obtain them. What sources are there for these data? Are they generally available to political researchers? Do the data exist at all? Are there problems of availability, reliability, and validity? Can these problems be successfully overcome? We discuss the principles of what makes good data in Chapter 7.

Data collection involves setting the boundaries for the study, and collecting information through observations, interviews, documents, visual materials, and published statistical and other data. You must set the boundaries for the study in advance of data gathering, including (a) the time frame, or *temporal domain*; (b) the place, or *spatial domain*; (c) the *actors* or units that are the relevant focus of the examination; and (d) the *variables* or factors you

think are important for arriving at an answer to your research question and that are, in fact, components of your hypothesis (answer). You are then ready to collect information. This information can be gathered from a variety of different sources.

There are many different ways to collect data: through ethnographic research, surveys and questionnaires, interviews, observations, or the analysis of existing documents and texts. Data can be gathered using more, or less, structured methods: for instance, using open-ended and flexible questions, or questions to be answered by respondents selecting from among a fixed set of choices. The main instrument for gathering data might be the researcher themselves (it may be that data are not 'collected' at all, but rather are co-produced by the observer and what is being observed), or instruments such as standardized surveys that are administered in a fashion that minimizes researcher bias. The method might entail gathering data through direct contact with the subject, as in the use of focus groups or ethnographic methods; or without direct contact, as in the case of mailed surveys. You can collect the data yourself through any of these means; use data collected by others, including statistical sources, oral histories, memoirs, newspapers, government documents, public opinion surveys, and interviews (see Chapter 10 'Using documents, archival sources, and historical writing as data'; and the Appendix); or use a combination of both. *Any data-collection method can be used for any type of research design*. No single source has a complete advantage over the others; and various combinations of methods might be complementary and used in tandem. Your hypothesis determines in large part the kind of data required and suggests methods for collecting it. However, practical limitations (e.g. time, money, skill, ethical concerns) will also enter into your calculations and choices.

Let's briefly consider some main data-collection methods, what each entails, and what sort of data or evidence each provides.

Questionnaires and surveys

Sometimes we need to know the personal experiences, perceptions, opinions, and attitudes of influential individuals in order to answer a research question. To get this information we can use questionnaires and surveys that ask people questions about particular topics that can reveal inside views of the political process. These can be mailed, handed out, or conducted in interview format (see Chapters 11 and 12).

Survey research is a method of gathering data from respondents thought to be representative of some population, using an instrument composed of closed or open-ended items (questions). It involves selecting the people to be approached (sampling), translating the broad objectives of the study into questions that will obtain the necessary information (questionnaire design), collecting data through questionnaires or interviews (fieldwork), and coding and inputting the responses (data processing). This is a major form of data collection in the social sciences, and is frequently used in political research.

Surveys combine a method of obtaining information from people by asking questions and modern random sampling procedures that allow a relatively small number of people to represent a much larger population (Schuman and Presser 1996: 1). They are a valuable resource for examining a wide range of topics, and can provide an accurate and reliable insight into what ordinary people think about politics and how they do politics. They rely

on samples, which are selected from a particular group of people or other units (such as households, businesses, schools, etc.) that the researcher wishes to study. If we were to interview everybody in a population (e.g. all voters, all asylum seekers), it would be incredibly time-consuming and expensive. Surveying a representative sample of the population allows the researcher to make generalizations about the attributes of a given population without having to actually interview everyone in that population. If the sample is chosen in a haphazard or subjective way, then there is little hope of making accurate generalizations about the wider population we are interested in.

Interviewing and focus groups

We have said that one of the ways researchers try to find out about the social world is to ask people questions, and that this can be done by asking people to fill in questionnaires (surveys). It can also be done through interviews conducted via telephone or internet, through formal face-to-face interviews, or more informally, in the context of focus groups (see Chapter 12).

Interviews can be structured or unstructured. In the structured interview, interviewees are asked a set of identical questions in exactly the same way, and are usually asked to select their answers from a limited range of options. More structured techniques might include surveys and questionnaires. Structured interviews are better for making comparisons and less structured interviews may be more appropriate for early exploratory phases of research. Unstructured interviews are more like ordinary conversations: there is no set interview structure and interviewees answer in their own words. These allow for longer questions and more in-depth probing, and are most frequently used in ethnographic research. The interviewer initiates the conversation, presents each topic by means of specific questions, and decides when the conversation on a topic has satisfied the research objectives. Researchers sometimes use a combination of interview methods, using structured questions to obtain factual information (such as age or income), and unstructured questions to probe deeper into people's experiences.

Focus groups

Focus groups involve a form of unstructured interviewing that generates different data from other forms of interviewing. A focus group is a group of people selected because they are believed to be related to some phenomenon of interest. The researcher meets with the group and facilitates an organized discussion related to something the participants have experience of, or beliefs about.

Discussion can bring out insights and understandings in ways which simple questionnaire items may not be able to tap into, such as emotional and unconscious motivations not amenable to the structured questions of conventional survey research. The interaction among focus group participants may reveal more than they would in the more formal interview setting. Using focus groups, researchers can learn how things are discussed in a particular culture, test hypotheses about beliefs, mix people who wouldn't normally mix, test questions for future interviews, or test policy ideas to see how citizens react to them.

Ethnographic research

Ethnographic research provides data on social phenomena by placing researchers 'in the midst of whatever it is they study so they can examine various phenomena as perceived by participants' (Berg 2004: 148). The objective of ethnographic research is to describe the lives of people other than ourselves with accuracy and detailed observation honed by first-hand experience. It involves participating, overtly or covertly, in people's daily lives in a way that can 'throw light on the issues that are the focus of research' (Hammersley and Atkinson 2007: 1). The idea is that a researcher must enter the environment of those under study to observe and understand their behaviour (see Chapter 13).

In political research, ethnographic studies have been conducted with the purpose of developing a better understanding of different institutions (e.g. the military, the BBC, the World Bank), or cultures (migrants in the US, fundamentalist religious groups in the EU), events (a war, a summit), or roles (campaign worker, news editor).

Participant observation involves the actual participation of the researcher in the events or environment being studied. This technique provides an 'insider' account of the activities and daily lives of the people being studied. By taking part in this way, researchers are able to construct an account of the way in which the group perceives their world. The personal insights of the researcher can be cross-validated through repeated, in-depth interviews with a broad cross-section of informants, and through conventional archival research, consultation with experts, use of surveys, and other techniques. The use of multiple researchers is an additional way to ensure reliability.

Discourse/content analysis

Content analysis generates data by analysing documents, reports, statistics, manuscripts, and other written, oral, or visual materials. These materials permit researchers to access subjects that may be difficult or impossible to obtain through direct personal contact (such as, for instance, interviews with decision-makers) or to increase the sample size above what would be possible through either interviews or direct observation (see Chapter 15).

Content analysis involves coding the 'content' of written documents, audio transcripts, radio programmes, television programmes, public speeches, or internet pages. It allows us to explore the beliefs, attitudes, and preferences of actors (even after they are dead).

Content analysis provides evidence about subjectivity: what were they (the actors) thinking? What were their intentions? Questions about subjective phenomena arise in virtually all types of social research. Narrative data (e.g. autobiographies, literature, journals, diaries, first-hand accounts, newspapers) often provide important keys to both process (and thus mechanisms) and subjectivity. These records provide different sorts of evidence. Verbal accounts from politicians, eyewitnesses, journalists, and contemporary historians constitute an important source of information for political research. These come increasingly with visual data via photographs, films, and videos. Participants in the political processes we study (e.g. civil servants, members of advisory councils, and representatives of pressure groups involved in decision-making processes) generate party programmes, parliamentary proceedings, resolutions, speeches, treaties, press conferences and press reports, television and radio interviews, and correspondence. Through content analysis, we can systematically

analyse these materials for clues to decision-makers' perceptions and attitudes. The method is often used in conjunction with other methods in order to establish a firmer causal link. For example, a researcher might explore whether a rise in racist statements in radio and television shows precedes the rise of racist attitudes in public opinion polls (using a survey).

Using data from existing archives allows you to widen the scope of your work far beyond what you could collect for yourself, and makes possible comparisons and the study of trends over time; and they can be used for purposes quite different from those for which they were originally collected. For instance, the Economic and Social Research Council (ESRC) Data Archive holds almost 5,000 data sets, including those from many large and important government-produced surveys and censuses, as well as academic research and historical materials. The Appendix contains a listing of data sources for political research, including databanks, consortia, data archives from state and independent government agencies, policy-making organizations, international organizations, academic institutions, scholars, think tanks, and private organizations.

It is also a good idea to use multiple sources of data and methods of data collection whenever possible. Doing this enables you to approach a research problem from different angles, something that is called 'triangulation'. Triangulation of evidence increases the reliability of the data and the process of gathering it. In the context of data collection, triangulation serves to corroborate the data gathered from other sources: the use of different data sources can enable researchers to cross-check findings. Triangulation yields more complete data and results in more credible findings; and it also enables researchers to find agreement between different perspectives. It might involve the use of different research methods to study a single research problem: a comparison of the results of these different methods can enable a researcher to identify whether differences are due to biases in one or another of these methods. Or it might involve the use of different researchers to study the same research problem with the same people and with the same methods: if the same results are discovered, the findings and interpretations have stronger validity because of this corroboration. Triangulation might also involve using different theoretical perspectives to look at the same data: examining the same data from different theoretical perspectives can enable the researcher to identify bias.

The Research in Action Box 6.1 provides an example of how rich data can be collected from multiple sources in order to address an issue of perennial interest to scholars in our field.

Ethical research

Ethical research is professional practice undertaken in accordance with a code of ethics—principles of conduct—for a given profession or group. All professions have a well-established code of ethics that governs the way they carry out research. The code varies from profession to profession, but there are certain behaviours in research—such as causing harm to individuals, breaching confidentiality, using information improperly, and introducing bias—that are considered unethical in any profession.

Ethical issues have become increasingly important in guiding research since the late 1990s or so. Before then, ethics was not something that researchers (not just ethnographers, but pretty much all political researchers) would have had to think very much about, as there

Research in Action 6.1 'The Paradox of Political Persistence'

In *Crucibles of Political Loyalty: Church Institutions and Electoral Continuity in Hungary* (Cambridge University Press, 2006), Jason Wittenberg seeks to explain the 'paradox of political persistence'—the intergenerational transmission of political preferences. 'Why,' he asks, 'do mass political loyalties persist long after the circumstances around which they arose have disappeared?' (2006: 1). Scholars have been very interested in this question; but they do not always link political attitudes to behaviour, or specify the mechanisms by which they are reproduced. Wittenberg uses a combination of *three different data-collection methods* to show (1) the durability of political beliefs, (2) their linkages to political behaviour, and (3) the processes by which these beliefs and behaviours maintain themselves through the passage of time.

1. Wittenberg focuses on a single *case study*. Hungary, he argues, represents a 'least likely' case of political persistence and so provides a tough empirical test of any argument advanced to explain it (we discuss types of case studies in Chapter 9). He explains that the communist regime that governed the country between 1948 and 1989 brought about a revolutionary transformation of society. But despite sweeping efforts to transform Hungarian society, and particularly rural strongholds of conservatism, the election of 1990, like that of 1948, resulted in a victory for the parties of 'the right'. Prior studies of political persistence have tended to focus on Latin America or Southern Europe and to attribute this outcome to the role of surviving civil society institutions. But because these institutions had been more extensively destroyed or co-opted in Hungary, the case potentially can reveal a new mechanism of transmission.

   So his question is: why, after forty years of communist rule, have *pre*-communist political cleavages manifested themselves again in *post*-communist Hungary? His answer focuses on the role of active local church organizations and church-based social networks which diligently and persistently sought to keep the conservative social milieu intact and, in spite of the communist regime's secularizing efforts were able to shape the voting preferences of the Hungarian electorate. This was the mechanism of transmission which enabled pre-communist loyalties to right-wing parties to survive communist rule and reassert themselves following its collapse.
2. *Statistical analysis* of electoral returns in some 3,000 municipalities across Hungary reveals the persistence of partisan loyalties throughout communist rule, with many localities showing support for precisely the same type of parties in 1990 that they had supported forty-two years earlier. It also establishes a link between patterns of electoral continuity and discontinuity and religiousity.
3. *Qualitative analysis* of party and church records in two sample counties, one predominantly Catholic and one predominantly Calvinist, then helps to expose the social mechanisms that explain the persistence of these attachments. It reveals abundant evidence that church leaders engaged in a persistent struggle with party officials to resist the regime's secularizing pressures, and to retain the loyalty, and shape the voting preferences, of the local populace. It shows that it was in those places that churches were able to retain their influence over the local population that right-wing voting in Hungary re-emerged after a half-century of communist rule. The longer a district was able, with the help of its local clergy, to resist secularization, the more likely it was to support the parties of the right in post-communist elections.

was little in the way of ethical scrutiny of research. But this has changed. Ethical scrutiny is now something that nearly all research is required to go through, even for graduate and undergraduate research projects. Most universities have ethics committees that issue guidelines about ethical practice, and these guidelines or codes are often based on the codes of professional organizations such as the British Sociological Association or the American Sociological Association.

Although one of the main goals of research is to accumulate knowledge or develop an understanding of a particular phenomenon, it is also recognized that achieving these goals should not come at the cost of all else. And in particular, the social cost (or potential cost) to those involved in the research (the informants) needs to be taken into account in order to ensure that they are not unduly exploited or harmed. Ethical scrutiny is therefore designed to protect research participants. It is also designed to try and protect the university institutions themselves, so that researchers are deterred from engaging in ethically dubious research activities that could come back to haunt the institution in the form of legal action or unwanted publicity.

We can therefore think of ethical issues in research in relation to the various 'stakeholders' in research. These include research participants, researchers, and sponsoring organizations.

Ethical issues to consider concerning research participants

Research participants include all those with direct and indirect involvement in a research project, as well as those who are likely to be affected by the findings of the project. For instance, the participants would include those who participate in a focus group, or who are involved in collecting or providing information relating to the project; those from whom information is collected or those who are studied by a researcher also become participants of the study. There are many ethical issues to consider in relation to the participants of a research activity. Broadly speaking, ethical issues can be summarized under three main headings: seeking consent; maintaining confidentiality; and the possibility of causing harm to participants. We briefly examine these areas below (and we expand on them in Chapters 8, 10, 12, and 13).

Informed consent

It is unethical to collect information without the knowledge, willingness, and 'informed consent' of participants.

The principle of **informed consent** requires that the researcher tell participants about the purpose and likely benefits of the study, the process through which participants were selected, procedures that will be followed, and any risks for the subject.

The issue of informed consent is one of the most hotly debated ethical principles. The bulk of discussion focuses on what is called disguised or **covert participation**. This type of observation happens without the informants' knowledge, and they are thus not provided with an opportunity to express whether they would like to take part in the research or not. Most ethical codes of conduct now regard this type of research as unjustified. It is often seen as a type of infiltration or spying (Bulmer 1982), and is even thought to contravene the human rights of autonomy and dignity of participants (Hammersley and Atkinson 2007). For this reason, covert studies are rarely carried out any more. The defence that the ends can sometimes justify the means tends not to carry much weight.

The principle of informed consent not only requires that the researcher should make their identity known, but that they should also make the purpose of their research known, and provide informants with the opportunity to withdraw from the research process at

any stage. The Social Research Association (SRA) *Ethical Guidelines* (2003: 27) goes as far as to say:

> Inquiries involving human subjects should be based as far as practicable on the freely given informed consent of subjects … They should be aware of their entitlement to refuse at any stage for whatever reason and to withdraw data supplied. Information that would be likely to affect a subject's willingness to participate should not be deliberately withheld.

The issue of deception is one that requires careful thought. It might be reasonably argued that any deception in the conduct of research involving human subjects is inconsistent with informed consent and is unethical. But it also may be necessary for participants to remain uninformed about the hypotheses under investigation, so as to ensure that the results of the study will not be biased as a result of this knowledge (McDermott 2002: 41). This is an issue, in particular, for experimental research. If deception is involved, researchers should explain to participants

> what the deception was and why it was deemed necessary for the unbiased collection of the data. In particular, subjects should be reassured that all their data will be confidential and that the experimenter will obtain subjects' written permission before any of their information is shared publicly.
>
> (McDermott 2002: 41)

Informed consent is, first of all, voluntary and uncoerced. As part of the procedure for obtaining informed consent, the researcher must ensure that potential participants are 'aware of their right to refuse to participate; understand the extent to which confidentiality and anonymity will be maintained; and be reminded of their right to renegotiate consent during the research process' (Hoong 2005: 279).

The researcher should always stress that participation in a study is completely voluntary, and that a person who declines to participate will not incur any penalty. People should know that they are not required to participate in a study, and that they can withdraw from a study at any point without penalty (de Vaus 2001: 83).

Informed consent is not only voluntary and incoerced, prospective participants must be given 'sufficient information about the project and what their participation in it entails' so that they will 'be able to make a reasoned decision as to whether to participate' (Schinke and Gilchrist 1993: 83). A researcher provides prospective participants with 'sufficient information' by informing them of the following:

- the type of information that is being sought. Do you intend to ask sensitive and intrusive questions? If so, you must inform prospective participants that you intend to ask these questions;
- the reason this information is being sought;
- how this information will be used;
- how they will be expected to participate;
- how their participation will directly or indirectly affect them. Prospective participants must be informed of potential risks they may incur by taking part the research.

Two other issues relating to informed consent have to do, first, with *providing incentives*: offering incentives to prospective participants (as, for instance, some material inducement)

to share information with you is unethical. A second issue is obtaining *consent from marginalized or vulnerable populations*. There are conditions which might render people incompetent to make an informed decision as to whether to participate: working with people in crisis—for instance, refugees—raises complex ethical issues around consent (as well as the possibility of causing harm). Other conditions which impact the ability to give informed consent are in cases where people cannot speak the language in which the research is being carried out, elderly and infirm individuals, and children.

Maintaining confidentiality

A second area of ethical concern relates to the right to privacy of the informants. This is partly related to issues of informed consent, and provides an opportunity for the informant to decide what information they are prepared to make public. But it also refers to the identity of the informant and their right to anonymity. This is important in all areas of research, but is particularly important when there is a risk that information that is provided to the researcher might be used against the informant at a later date. Protecting the privacy and confidentiality of what informants tell you can therefore be a very serious issue. At one level, in ethnographic research, it is easy to use pseudonyms for the informants so that they cannot be identified by name. But this may not always be enough to ensure privacy. Informants may be identifiable by information you have revealed about them. Some ethnographers therefore prefer to make the field site anonymous.

Sharing information about a respondent with others for purposes other than research is unethical. You may need to identify your study population in order to put your findings into context; but it is unethical to identify an individual participant and the information he or she provided. You must protect the confidentiality and anonymity of the information gathered from your respondents.

The possibility of causing harm to participants

Another area of ethical concern relates to the harm of participants. Harm includes 'research that might involve such things as discomfort, anxiety, harassment, invasion of privacy, or demeaning or dehumanising procedures' (Bailey 1978: 384). Researchers must ensure that the risk of harm is minimal, that is, not greater than that ordinarily encountered in daily life.

Research that is likely to harm participants is generally regarded as unacceptable. But harm can occur in a variety of ways. Although political research doesn't tend to endanger informants directly, the research process can nonetheless have consequences for informants which may cause them harm or distress. Sometimes this can come about from the act of research. For example, research on racism might reinforce racist practices (see discussion by Troyna and Carrington 1989) or research on football hooligans might incite informants to violence as a way of impressing the researcher. But harm can also come from the publication of research, and the making public of details of people's lives or opinions, which may cause them distress or even put them in danger from the authorities. The identity of informants should be kept confidential; but more than that, it should not be possible to identify them.

Ethical issues to consider relating to the researcher

Areas of ethical concern with regard to the researcher include introducing bias; inaccurate reporting; and the inappropriate use of information.

Avoiding bias

It is unethical to introduce bias into a research activity. Bias is a deliberate attempt either to hide what you have found in your study, or to highlight something disproportionately to its true existence. We discussed the ways in which a researcher's biases can intrude on the research process in Chapter 3. We consider further problems of bias, misrepresentation, or the selective presentation of evidence in Chapter 10.

Incorrect reporting

To report the findings in a way that changes or slants them to serve your own or someone else's interest is unethical. Correct and unbiased reporting of the findings are important characteristics of ethical research practice.

Inappropriate use of the information

The use of information in a way that directly or indirectly affects participants adversely is unethical.

Ethical issues regarding the sponsoring organization

Funding organizations responsible for financing a research activity are also stakeholders in research. These might be government or academic institutions, or research bodies which provide funds for specific purposes. Ethical considerations in relation to sponsoring organizations concern restrictions imposed on research designs and the possible use of findings.

Restrictions imposed by the sponsoring organization

Most research in the social sciences is carried out using funds provided by sponsoring organizations for a specific purpose. The funds may be given to study the impact of a policy, or the behaviour of a group or community; or to study a phenomenon, issue, or attitude. Sometimes there may be direct or indirect controls exercised by sponsoring organizations. They may select the methodology, or impose other restrictions on the research that may stand in the way of obtaining and disseminating accurate information. Both the imposition and acceptance of these controls and restrictions are unethical, as they constitute interference and could amount to the sponsoring organization tailoring research findings to meet its vested interests.

The misuse of information

How is the sponsoring body going to use the information? How is this likely to affect the study population?

All of these issues are discussed in more detail in the Social Research Council's online publication, *Ethical Guidelines* (posted at http://the-sra.org.uk/wp-content/uploads/ethics03.pdf).

The ethical issues we have just reviewed will not be equally relevant to all the approaches we discuss in the remaining chapters of this book. Considerations relating to voluntary participation, informed consent, privacy, harm, and exploitation will have most relevance to experimental research, surveys, interviewing and focus groups, and ethnography and participant observation. They will have least relevance for unobtrusive forms of research such as historical research and content analysis, both of which involve the examination of written documents and analyses of existing sources. Of course, all types of research must meet standards of good research practice and scientific integrity, something which we emphasize throughout this book and in relation to each approach we discuss. Guidelines for our discipline have been established by the American Political Science Association's *Guide to Professional Ethics in Political Science* (2012). The *Guide* states that 'researchers have an ethical obligation to facilitate the evaluation of their evidence-based knowledge claims through data access, production transparency, and analytic transparency so that their work can be tested or replicated'. The definition of these terms, as provided in the *Guide*, is shown in Box 6.3.

BOX 6.3 Ethical Obligations of Researchers

DATA ACCESS: Researchers making evidence-based knowledge claims should reference the data they used to make those claims. If these are data they generated or collected themselves, researchers should provide access to those data or explain why they cannot.

PRODUCTION TRANSPARENCY: Researchers providing access to data they generated or collected themselves should offer a full account of the procedures used to collect or generate the data.

ANALYTIC TRANSPARENCY: Researchers making evidence-based knowledge claims should provide a full account of how they draw their analytic conclusions from the data, that is, clearly explicate the links connecting data to conclusions.

Conclusions

Perhaps more than any other academic discipline, political research incorporates a wide variety of different methods and approaches. This has both advantages and disadvantages. It strengthens the discipline as a whole when this diversity is embraced, and when researchers adopt and integrate the different approaches and engage with research from across the methodological spectrum. It weakens the discipline when this diversity becomes segregated, and when researchers from different methodological traditions retreat into their own enclaves and do not engage with what other people are doing in the discipline. In the following chapters we present chapter-length treatments of some of the most widely used and influential methods and approaches in the study of politics. Each method has its own strengths and weaknesses, but rather than getting bogged down in petty discussions about which method is best, or which method is best for which type of topic or question, the

challenge that faces serious researchers is to think about how these methods can be incorporated with each other, and the strengths of one approach used to balance the weaknesses of another.

Questions

1. What is meant by 'research design'?
2. What are the various types of research designs used in political research?
3. Why should researchers be concerned with research design? Can politics be studied rigorously without attention to issues of research design?
4. What are the major components of research designs?
5. What are the general criteria for evaluating research designs?
6. What are the various potential stakeholder groups in a given research project? What are the possible ethical concerns you need to be aware of from the perspective of each one of these groups?

Guide to Further Reading

Creswell, J. W. (2003), *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (Thousand Oaks, CA: Sage Publications).
An introduction to different types of research design, and how to combine qualitative and quantitative research strategies.

Davies, R. B. (1994), 'From Cross-Sectional to Longitudinal Analysis', in A. Dale and R. B. Davies (eds), *Analyzing Social and Political Change: A Casebook of Methods* (Thousand Oaks, CA: Sage Publications).
Davies argues that more data are required to characterize empirically the dynamic process that lies behind the cross-sectional snapshot; and that these data can be supplied through longitudinal analysis.

de Vaus, D. (2001), *Research Design in Social Research* (Thousand Oaks, CA: Sage Publications).
De Vaus presents key types of social science research design, including case studies, cross-sectional, experimental, and longitudinal studies, with a discussion, for each, of tools required, possible issues, and data analysis.

Druckman, J. N., D. P. Green, J. H. Kuklinski, and A. Lupia (2006), 'The Growth and Development of Experimental Research in Political Science', *American Political Science Review* 100(4): 627–35.
This article documents how thinking about experimentation has evolved over the century, and demonstrates the growing influence of laboratory, survey, and field experiments.

Flyvbjerg, B. (2006), 'Five Misunderstandings about Case-Study Research', *Qualitative Inquiry* 12(2) (April): 219–45.
This article discusses five misunderstandings of case study research: (a) theoretical knowledge is more valuable than practical knowledge; (b) one cannot generalize from a single case; therefore, the single-case study cannot contribute to scientific development; (c) the case study is most useful for generating hypotheses, whereas other methods are more suitable for hypothesis testing and theory building; (d) the case study contains a bias towards verification; and (e) it is often difficult to summarize specific case studies.

George, A. L. and A. Bennett (2005), *Case Studies and Theory Development in the Social Sciences* (Cambridge, MA: MIT Press).
This book discusses the use of case studies to build and test theories in political science and the other social sciences, and presents a comprehensive analysis of research methods using case studies.

Gerring, J. (2001), *Social Science Methodology: A Criterial Framework* (Cambridge: Cambridge University Press), 'Research Design: General Criteria', chapter 8; 'Methods', chapter 9; and 'Strategies of Research Design', chapter 10.
This offers a clear and practical guide to the development of qualitative and quantitative research designs, as well as a useful perspective on how we should conceptualize methodological problems.

Gerring, J. (2004), 'What is a Case Study and What is it Good For?', *American Political Science Review* 98 (May): 341–54.
This article clarifies the meaning, and explains the utility, of the case study method. It argues for the complementarity of single-unit and cross-unit research designs.

McDermott, R. (2002), 'Experimental Methods in Political Science', *Annual Review of Political Science* 5: 31–61.
This article discusses issues of central concern to experimentalists, including: impact vs. control, mundane vs. experimental realism, internal vs. external validity, deception and laboratory vs. field experiments; it also summarizes the advantages and disadvantages of experimentation.

Munck, G. (1998), 'Canons of Research Design in Qualitative Analysis', *Studies in Comparative International Development* 33(3) (September): 18–45.
This article dvances the idea of a multi-method approach to political research.

Ragin, C. (2000), *Fuzzy-Set Social Science* (Chicago: University of Chicago Press).
Explains the difference between research designs aimed at testing theory and those aimed at building, developing, and refining theory.

Wittenberg, J. (2007), 'Peril and Promise: Multi-Method Research in Practice', *Qualitative Mixed Methods Newsletter* 5(1): 19–22.
This is an essay, addressed to graduate but also undergraduate dissertation writers, in which the author attempts to 'demystify' the use of multi-method research' through illustrating how he executed the project that resulted in his dissertation (and ultimately the book based on it, *Crucibles of Political Loyalty: Church Institutions and Electoral Continuity in Hungary* (Cambridge University Press, 2006)).

Ethical Research: Additional Resources and Readings

Chatty, D., G. Crivello, and G. Lewando Hundt (2005), 'Theoretical and Methodological Challenges of Studying Refugee Children in the Middle East and North Africa: Young Palestinian, Afghan and Sahrawi Refugees', *Journal of Refugee Studies* 18(4) (2005): 387–409.

Hart, J. and B. Tyrer (2006), 'Research with Children Living in Situations of Armed Conflict: Concepts, Ethics and Methods', Refugee Studies Centre Working Paper 30. Available at: http://www.rsc.ox.ac.uk/PDFs/workingpaper30.pdf

Jacobsen, K. and L. Landau (2003), 'Researching Refugees: Some Methodological and Ethical Considerations in Social Science and Forced Migration', *New Issues in Refugee Research* 90. Available to download at: http://www.unhcr.org/research/RESEARCH/3f13bb967.pdf

Pittaway, E., L. Bartolomei, and R. Hugman (2010), '"Stop Stealing Our Stories": The Ethics of Research with Vulnerable Groups', *Journal of Human Rights Practice* 2(2) (June): 229–51.

Pitts, M. and A. Smith (eds) (2007), *Researching the Margins: Strategies for Ethical and Rigorous Research with Marginalised Communities*. New York: Palgrave MacMillan.

Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. Available to download at: http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/Default/

Wood, E. (2006), 'The Ethical Challenges of Field Research in Conflict Zones', *Qualitative Sociology* 29(3) (September): 373–86.

References

Allison, G. T. (1969), 'Conceptual Models and the Cuban Missile Crisis', *American Political Science Review* 53: 689–718.

American Political Science Association (2012), *A Guide to Professional Ethics in Political Science*, 2nd revised edition (Washington, DC: American Political Science Association).

Bailey, K. D. (1978), *Methodes of Social Research*, 3rd edition (New York: Free Press).

Berg, B. L. (2004), *Qualitative Research Methods for the Social Sciences* (Boston, MA: Allyn and Bacon).

Bulmer, M. (1982), 'The Merits and Demerits of Covert Participant Observation', in M. Bulmer, (ed.), *Social Research Ethics* (Basingstoke: Macmillan), 217–51.

Collier, D. and J. Mahoney (1996), 'Research Note. Insights and Pitfalls: Selection Bias in Qualitative Research', *World Politics* 49(1): 56–91.

de Vaus, D. (2001), *Research Design in Social Science* (London: Sage).

Gerring, J. (2007), *Case Study Research: Principles and Practices* (Cambridge: Cambridge University Press).

Green, D. and A. Gerber (2003), 'The Underprovision of Experiments in Political Science', *Annals of the American Academy of Political and Social Science* 589: 94–112.

Hammersley, M. and P. Atkinson (2007), *Ethnography: Principles in Practice*, 3rd edition (London: Routledge).

Hoong Sin, C. (2005) 'Seeking Informed Consent: Reflections on Research Practice', *Sociology* 39(2): 277–94.

King, G., R. Keohane, and S. Verba (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ: Princeton University Press).

Leighly, J. (1995), 'Attitudes, Opportunities and Incentives: A Review Essay on Political Participation', *Political Research Quarterly* 48 (March): 181–209.

McDermott, R. (2002), 'Experimental Methods in Political Science', *Annual Review of Political Science* 5 (June): 31–61.

Mishler, W. and R. Rose (2007), 'Generation, Age and Time: The Dynamics of Political Learning during Russia's Transformation', *American Journal of Political Science* 51(4) (October): 822–34.

Morton, R. B. and K. Williams (2010), *From Nature to the Lab: The Methodology of Experimental Political Science and the Study of Causality* (Cambridge: Cambridge University Press).

Ostrom, E. (2007), 'Why Do We Need Laboratory Experiments in Political Science?' Paper presented at the 2007 American Political Science Association annual meeting, Chicago, IL, 30 August–2 September.

Pierson, P. (1994), *Dismantling the Welfare State?* (Cambridge: Cambridge University Press).

Rawls, J. (1993), *Political Liberalism* (New York: Columbia University Press).

Roberts, T. J., B. C. Parks, and A. A. Vasquez (2004), 'Who Ratifies Environmental Treaties and Why? Institutionalism, Structuralism and Participation by 192 Nations in 22 Treaties', *Global Environmental Politics* 4: 22–64.

Schinke, S. P. and L. Gilchrist (1993), 'Ethics in Research', in R. M. Grinnell (ed.), *Social Work, Research and Evaluation: Foundations of Evidence-Based Practice*, 4th edition (Itasca, IL: F. E. Peacock).

Schuman, H. and S. Presser (1996), *Questions and Answers in Attitude Surveys* (Beverly Hills, CA: Sage).

Shively, W. P. (1989), *The Craft of Political Research*, 2nd edition (Englewood Cliffs, NJ: Prentice-Hall, Inc.).

Social Research Association (SRA) (2003), *Ethical Guidelines* (London: Social Research Association). Available at: http://the-sra.org.uk/wp-content/uploads/ethics03.pdf

Troyna, B. and B. Carrington (1989), 'Whose Side are We On?', in R. G. Burgess, (ed.), *The Ethics of Educational Research* (Lewes: Falmer Press), 205–23.

White, L. and R. P. Clark (1983), *Political Analysis: Technique and Practice* (Monterey, CA: Brooks/Cole Pub. Co.).

Yin, R. (1984), *Case Study Research: Design and Methods*, 2nd edition (Thousand Oaks, CA: Sage Publications).

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

Endnotes

1. Of course, it wouldn't hurt for the student in this example to take advantage of an opportunity to interview government officials in Egypt, as it might provide useful insights and perhaps generate an idea for a research question. But we would, nonetheless, use the query as an opportunity to make the point that research questions and hypotheses, not methods, should drive research.
2. We made the point, in Chapter 4, that research should be driven, not by methods, but by questions and problems. But, to continue with the toolbox metaphor, there are researchers who will start the research process by selecting their preferred tool from the research methods toolbox and then go looking for something to use it on.
3. It cannot be scientifically demonstrated that certain normative standpoints are more correct (in the sense of closer to the truth) than others. Normative questions typically require research that combines given political facts with moral arguments. In contrast to empirical types of research, the researcher is not 'required to produce the full factual basis for his/her argument'. It is often the case that the researcher will 'need to assume facts which cannot possibly be tested against reality' (Shively 1989: 9).

What is Data?

Chapter Summary

Previous chapters focused on how to find and formulate a research question (Chapter 4), how to develop a hunch about the answer to one (Chapter 5), and how to design research in order to investigate your hunch (Chapter 6). In the following chapters we discuss different methods of data collection and analysis. In this chapter, we focus on data. Our discussion focuses on the following issues:

- from concepts to measurement;
- validity;
- reliability;
- selection issues;
- types of data;
- big data;
- data quality checklist.

Introduction

If we want to develop our understanding about the political world or investigate different ideas or explanations about how things work, we need to think about how our ideas relate to what happens in practice. This means engaging with the empirical world—and to do this we need to collect information. Our record of the information that we have collected is data. We then use this data as evidence to support our arguments. Data thus constitutes the information base for our analysis. Without data—or evidence—argument is little more than conjecture.

In the following chapters, we will discuss different methods of data collection and analysis. However, before doing that, the purpose of this chapter is to outline what we mean by data in the first place. In this chapter, we discuss the main types of data used in Politics and International Relations, and the main criteria by which we can judge whether our data is good, or not. Whatever type of data we use in our analysis, there are a number of issues we must confront. First, we need to ensure as far as possible that our data is valid. That is, the data does actually measure what we claim it does. And second, we need to ensure that our data is reliable. That is, we have recorded the data (or information) accurately. All observation and measurement in the social sciences is imprecise. We therefore

need to take these issues very seriously. Measurement error can be thought of as the difference between the true value of our variable of interest and our recording of that value. Our data is never perfect—and so we must take steps to ensure that it is as good as possible. Good data helps us to make good inferences and draw strong conclusions. If the data is bad, then no matter how sophisticated our analysis, our conclusions will be found wanting.

From concepts to measurement

Research in political science usually starts with a hunch or a tentative theory about some aspect of political reality. To systematically investigate these hunches or ideas about different political phenomena it is necessary to move from the theoretical or abstract world into the empirical world—and this means thinking about what data or evidence is relevant to answering your question (see Figure 7.1). There are a number of steps involved in this process. First, we must have a clear idea of what it is that we want to analyse. Suppose we are interested in analysing democracy—and whether a specific country is democratic or not (or has become more or less democratic over time). Before we move to data collection, it is important that we have a clear *conceptual definition* of what it is that we want to investigate, and what it is that we mean by the term 'democracy'. This process of deciding what we mean by a specific term is called conceptualization.

Democracy is a concept where there is considerable disagreement over what it means. Indeed, Michael Coppedge (2012: 11) refers to democracy as an *essentially contested* concept because although nearly everyone values the label, there are 'different reasonable and legitimate, yet incompatible, criteria for judging whether the label is deserved'. To some people, democracy is simply a way of selecting a government; but to others it encompasses

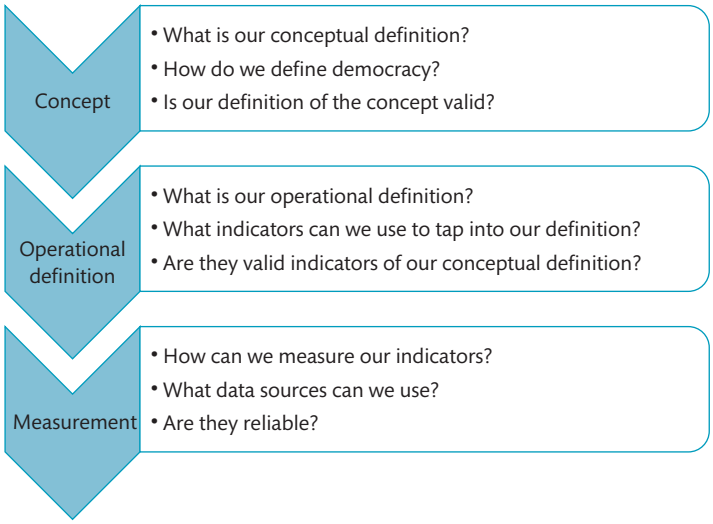


Figure 7.1 From concepts to measurement

a much broader array of social and political rights. No definition of democracy will please everyone, but unless we want to get bogged down in semantic confusion we must nonetheless be clear about what we mean by the term.

Let's say that we want to adopt a minimalist definition of democracy and, following scholars like Vanhanen (1997, 2000), we define democracy as a system of electing governments through competitive elections, with full adult franchise. This definition emphasizes two central tenants of democracy to do with competition and participation.

The first challenge we face is to justify whether our conceptual definition is valid or not. There may be widespread disagreement about how to define a concept, so it is important to be clear at the outset what is included in our definition and what is excluded, and to provide a justification for the course of action we have taken. For example, we may acknowledge that there is more to democracy than simply holding elections, but that it is also true that democracy cannot exist without elections—and so, for the purposes of our research, we decide to focus on the minimum institutional conditions that must be met for democracy to exist. And, while not all countries that meet this minimum definition may be regarded as fully fledged democracies, clearly all those which fail to meet this condition can be regarded as non-democracies.

Having established a clear conceptual definition the next task is to think of an *operational definition*. Our conceptualization of democracy emphasizes competition and participation, and so we must think about how we can operationalize these terms. The process of operationalization involves moving from the abstract to the specific. Thus we must provide an operational definition of our two main components of democracy. For example, we may say that our operational definition of competition is the difference in vote share between the winning party in an election and the second placed party in that election. So, when one party wins by a huge margin we can say that elections are not very competitive, but when one party wins by just a few votes we can say that elections are competitive. For instance, Saddam Hussein used to regularly hold elections in Iraq, where he was reported to have received more than 99 per cent of the vote. According to our definition of democracy, this would be an example of an 'uncompetitive election'.

Turning to our second indicator of democracy, our operational definition of participation might be the proportion of the adult population that turns out to vote. Thus, in countries where there is full adult franchise, and everyone is registered to vote, and then turns out to vote in an election, participation will be high. But in countries where some groups are not eligible to vote (such as women, or different ethnic or racial groups); or where there are barriers to registration (such as literacy tests); or where turnout is low; then participation will be low. For instance, during apartheid in South Africa only white people were allowed to vote and the black majority were barred from participating. According to our definition of democracy, this would be an example of restricted participation.

The main issue to consider with operational definitions is whether or not they are valid indicators of our concept. That is, once we have defined democracy by competition and participation, is the margin of victory in an election a valid indicator of competition? And is the level of adult participation in an election a valid indicator of participation? We may object to democracy being defined in such narrow terms, but concede that the operational indicators are valid. In this respect it is important to be internally consistent and to make sure that you do try and measure what you say you are going to study.

The last step in the process is to think about how we can actually measure our **indicators** of interest and what data we can collect. To measure competition, we may decide to simply inspect official election returns that are published after all the votes have been counted. Similarly, to measure participation we may count the number of votes that have been cast in the election and compare this to the most recent census estimates of adult population size in the country. The key issue here is to do with reliability. Are the election results an accurate reflection of what actually happened in the election? Maybe they have been manipulated in some way. Perhaps there has been electoral fraud. Similarly, does the country have reliable population estimates? Are there existing sources that you can use?

Validity

Generally speaking, whatever type of data is collected or used, the researcher will have to confront issues of **validity** and reliability. If the data itself is not valid or reliable, then the insights that are garnered from analysing the data are likely to be misleading in one way or another. Continuing with the example of democracy, the issue of validity is not whether we have measured participation and competition properly but whether these are suitable indicators of our concept—democracy. The validity of a measure therefore depends upon how we define the concept. This is always going to be open to debate, though it is obviously more straightforward in some cases than others. There are no easy ways to resolve this issue. But there are a number of basic ways in which we can try and assess the validity of our indicators (see Moser and Kalton 1971 and de Vaus 2002 for an extended discussion) and defend the approach we have taken.

The first and most straightforward is **face validity**. Face validity simply means that, on the face of it, the indicator intuitively seems like a good measure of the concept. Thus we may all agree that participation is an important element of democracy and so has a certain amount of face validity as an indicator; but if we were to suggest that we could measure democracy by looking at incarceration rates we might question what, if anything, this has to do with our central concept of interest. Face validity thus refers to whether there is broad agreement that the indicator is directly relevant to the concept. If the answer is 'no' then we definitely have a problem, but what is intuitively 'yes' for one person may not be so for another.

Content validity examines the extent to which the indicator covers the full range of the concept, covering each of its different aspects. Thus, we might argue that the indicators of democracy lack content validity as democracy encompasses more than just participation and competition. Our measure therefore neglects to take into account important political rights, such as the right to self-expression, the right to seek out alternative forms of information, and the right to form independent associations and organizations, which scholars such as Robert Dahl (1989) have suggested are important preconditions for the presence of democracy. When only one or two dimensions of a multifaceted concept are measured, it is important to be explicit about which dimensions have been left out (and why), as well as which dimensions have been included. However, in general it is good practice to avoid over-claiming what your indicators measure. So if you are just measuring political competition, it is better to be explicit about this at the beginning—and say that your project

is concerned with analysing why some countries are more competitive than others, rather than why some countries are more democratic than others.

A third way to evaluate the validity of an indicator is by empirical analysis. **Construct validity** examines how well the measure conforms to our theoretical expectations by examining the extent to which it is associated with other theoretically relevant factors. So, on theoretical grounds we could assert that democracy should be associated with freedom of the press. Thus, we might not regard a free press as being part of our definition of democracy, but nonetheless expect democracies to respect press freedoms more than non-democracies. And so, in order to validate our measure of democracy we would see whether it is associated with press freedom. The presence of such a relationship lends some support to the validity of the measure. Ultimately, though, the matter of validity is one of judgement.

When we deal with difficult-to-define or multifaceted concepts, these issues are difficult to deal with. It is therefore sometimes tempting to skirt around them and avoid making precise statements about what we are doing, or why. However, this approach leads to very vague and impressionistic conclusions, which lack a strong empirical core. When a concept is difficult to define it is more important than ever to be explicit about how we are going to operationalize it. It is only when we have a clear idea of what it is we want to find out that we can begin to think about what sort of data might be useful. Even then, it is sometimes very difficult to identify what type of data we should use, and whether there is existing data we can analyse or new data we can collect.

Example: measuring corruption

We can illustrate some of these issues with recent attempts to measure corruption. Corruption is a major topic of interest in Politics and International Relations, but it is also a difficult topic to analyse because, by definition, corruption is a clandestine activity that is difficult to directly observe. Any study that wishes to investigate corruption therefore has to confront a series of difficult questions about how to operationalize and measure the incidence of corruption in different parts of the world (for a review, see Heath, Richards, and de Graaf 2016).

The first step is to be explicit about what is meant by corruption. A simple and widely-used definition, used by Transparency International, the World Bank, and the US Department of Justice, is that corruption is 'the misuse of public position for private gain'. Typically, this involves paying a government official a bribe in order to receive goods or a service. The problem is—because of the clandestine nature of the activity—it is very challenging to think how we could actually measure this type of activity. Obviously we could just ask public officials whether they ever demand bribes, but it is doubtful whether they would reply honestly. Instead, researchers have concentrated on trying to develop a number of indirect measures of corruption (see Figure 7.2).

One widely used approach is to rank countries according to their level of 'perceived' corruption. For example, Transparency International regularly carries out a survey which asks business people to rate the level of corruption in different countries. Thus, the operational definition of corruption is not strictly the incidence of corruption but the perception of corruption among a group of people who should be well-placed to judge the business practices

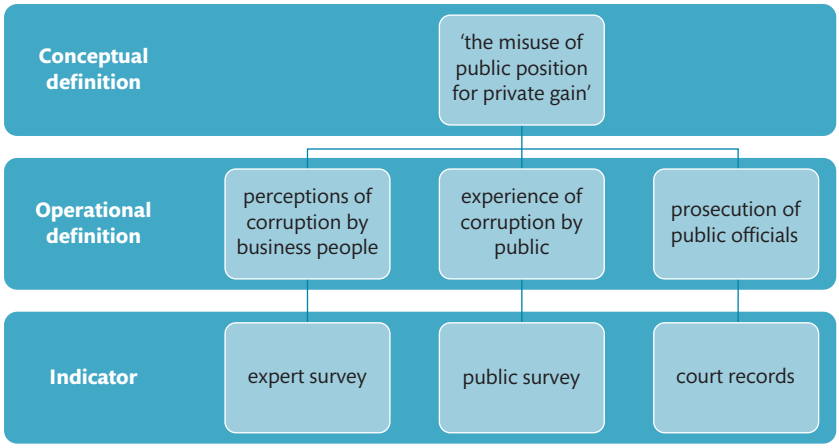


Figure 7.2 Approaches to measuring corruption

of different countries around the world. This kind of 'perception' data is widely used in Politics and International Relations. However, a major issue with this type of data is whether perceptions of corruption are a valid indicator of the actual incidence of corruption. Does the perception that a country is corrupt mean that a country really is corrupt? People's perception of the level of corruption in a country might be shaped by all sorts of factors—such as their political ideology, exposure to media, or even whether they like the country or not and enjoy spending time there—so the indicator may not be particularly valid.

So, what other types of data could therefore provide a better source of information? Some scholars have used conviction statistics to measure how many public officials are successfully prosecuted on charges of corruption (e.g. Glaeser and Saks 2006 on Federal prosecutions in the USA). But this indicator is also problematic as they may just reflect the energy and commitment of law enforcement agencies rather than the actual incidence of corruption (Goel and Nelson 2011). Lastly, other scholars have used 'victim' surveys to measure whether or not members of the public have actually had to pay a bribe in order to receive a good or service. This issue of validity remains with victim data, though in many respects they are the least problematic of the measures currently in use (Heath, Richards, and de Graaf 2016).

Reliability

Related to the issue of validity is the issue of reliability. Broadly speaking, reliability refers to how accurately we have measured our indicator. Thus, even if we have a valid indicator, if we are not able to measure it accurately we may end up with unreliable data. In more technical terms reliability means the measurement of an indicator yields the same results on repeated occasions. The higher the reliability of the measuring procedure the more consistent the results will be. Data can be valid but not reliable, and can be reliable but not valid. Obviously we would prefer it if our data was both valid and reliable.

There are a number of different ways of assessing the reliability of our data—and these techniques vary somewhat according to the type of data we are interested in. To illustrate some of the more common ways of doing this let's continue with the example of measuring corruption. We may have concerns about the validity of using perceptions of corruption to measure actual corruption; but should we also have concerns about the reliability of this approach? If the data is reliable then there should be broad agreement between business people about which countries are corrupt and which countries are not. If there is a lot of disagreement between business people about a country's level of corruption—and some people, for instance, regard business practice in India as very corrupt, whereas others do not—then our data may be unreliable. We refer to this type of reliability as **intercoder reliability** which is discussed in more detail in Chapter 14. Intercoder reliability reveals the extent to which different coders (in this case business people), each coding the same content (a country's level of corruption), come to the same coding decisions (whether the country is corrupt or not).

Another way in which researchers often talk about reliability is whether our measurement of our indicators produces consistent results on repeated occasions. This is somewhat different to intercoder reliability and is generally used in different contexts. So, for example, if we wanted to know whether the 'victim' surveys produced reliable data on whether people had experienced paying a bribe to a government official we might want to ask our respondents questions about this on more than one occasion. This is sometimes known as the *test-retest* method. It involves applying the same test to the same observations at different moments in time and comparing the results. We discuss this in more detail in Chapter 11. If all the tests produce the same result we have a reliable measure—but if they do not then we don't. It is clear that unreliable data lacks validity. But reliable data need not be valid, because it could be measuring something other than what it was designed to measure, and high reliability should not compensate for low validity.

How to choose cases

In addition to these issues, all data need to confront the issue of *external validity*. This means that we have to think about how far we can generalize the findings from our data. Sometimes people will make grand generalizations, say that they have discovered some universal truth, and that their findings apply across time and space. However, more often we will make bounded generalizations and say that our findings apply to a particular country at a particular point in time; or perhaps even to just a particular group within that country. Making generalizations doesn't mean that your findings apply to everyone (or everything) but that your findings apply to more than just the people you have spoken to or the particular cases you have collected data on. If we are unable to make any generalizations—or wider inferences—then our data is little more than an anecdote.

There are often thought to be two trade-offs with respect to data collection. On the one hand we want our data to be internally valid and to measure what we are interested in. So, to take another example, if we want to know what someone thinks about immigration, we might think it important to have a detailed conversation with them; talk to them about their experiences, perceptions, and perhaps even observe their interactions. But, we often want to draw conclusions as well, and in this respect it is important that our data is externally valid.

While it is interesting to know what our neighbour thinks about immigrants, we might also wonder how typical our neighbour is and whether their views are also shared by others. This brings us to how far we can make generalizations from the data we have collected.

In order to set the parameters for generalization, it is important to think about case selection. This is a critical, though often underappreciated part of data collection, and to a large extent determines what is good data (or at least useful data) from bad data. When we collect data we often deal with samples, however big or small those samples are. For example, if we are doing a study on young people, we do not interview all the young people in the world, or even in one particular country. Rather, we select a sample to interview. This sample might be a few thousand people if we are conducting a survey—it might be just a handful of people if we are conducting a semi-structured interview. The people that we end up interviewing represent our cases, and for each case we collect data.

The issue of case selection is extremely important because the cases we choose to analyse and collect data on can influence the answers we get to a particular question. When this occurs, it is known as **selection bias**. If we do not think carefully about case selection we can end up with answers that are not very robust, or that are even misleading. That is, we will make faulty inferences. The best way to avoid this happening is to conduct a large-N study and select cases using random probability methods, or to take a full census of cases (see Chapter 11). But this is not always a viable option. To a certain extent then, selection bias, or at least the risk of selection bias, is always going to be a danger in small-N studies, particularly when the study is designed to try and make some sort of generalization about a wider population, however tentatively put. We discuss the issue in more detail in Chapters 9 and 11.

In order to minimize these threats to external validity it is important to follow a number of steps (see Figure 7.3). Suppose you were interested in examining when the UN Security Council intervenes in domestic conflicts. You might have a range of hypotheses about what motivates the UN to intervene in some conflicts but not others, from humanitarian need to political expediency (see Binder 2015). To investigate your hypotheses you

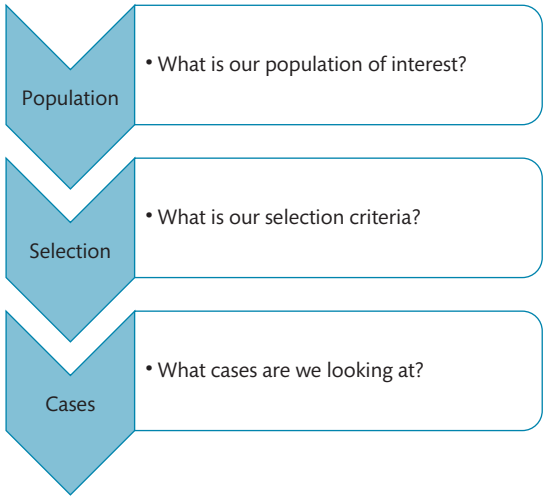


Figure 7.3 From populations to cases

will need to choose some cases to look at and collect data on; but which cases will you look at and how will you choose them? Sometimes students will rush to focus on one or two high-profile cases without considering the broader range of cases they could potentially study. The first step then is to define your population of interest. The population in this context refers to all the possible cases you could theoretically pick to analyse. Since a case is a domestic conflict, the population for your study would be all the domestic conflicts that have taken place in the world since, say, 1946. This is obviously going to involve a lot of cases, and it would be a lot of work to find out how many conflicts there have been from scratch. The Peace Research Institute Oslo (PRIO) keeps a detailed log of all such conflicts.

The next step is to outline a clear rationale for deciding which particular case or cases you are going to look at. The key point is that there must be some rationale for selecting some cases, but not others—and this rationale should be related to the theories or ideas that you want to investigate. There are various ways of doing this which we discuss in more detail in Chapter 9—but unless there is a strong justification for case selection it is very difficult to know what the broader implications of your research are likely to be. For instance, you could decide to pick a few cases where there has been a clear humanitarian crisis, and use this as a basis for investigating whether political considerations enter in the calculus of whether or not the UN decide to intervene. Having decided upon selection criteria, the last step is to pick the cases that you will actually investigate. Following this approach allows you to contextualize your data and draw stronger conclusions.

Although this approach to case selection sounds quantitative, the way in which we choose a sample depends more upon the type of information we want to find out and the nature of the conclusion we want to draw than it does on the type of method that we use. Large samples are not always necessary, or even desirable, and valid conclusions can still be drawn from analysing just a few cases, but only for particular types of research questions. For example, suppose you wanted to find out about the policy-making process in a particular country. You could approach this topic in a number of different ways depending upon what you wanted to find out. First, suppose you wanted to find out how the policy-making process works. To do this, you might want to interview 'experts' who have first-hand knowledge about the policy-making process (e.g. policy-makers themselves). You could just interview one policy-maker and ask them how the process works. Hopefully, the expert you have chosen to interview is competent and can give you a reasonably accurate account of the process. However, just to be on the safe side you might want to interview more than one policy-maker. If you interview five policy-makers and they all give you a similar account about how the policy-making process works, you can be reasonably confident in the data you have collected. Thus, you can make a valid inference—or a valid conclusion about your object of interest (the policy-making process) from just a handful of interviews. But suppose you had a slightly different research objective and wanted to know what policy-makers *thought* about the policy-making process and whether they thought it worked well or not. Now you would want to make a very different sort of inference—not about how something works, but about what people think, and this means you would want to draw a conclusion about a broader population of interest (policy-makers). In this situation you would be very unwise to draw strong conclusions about what policy-makers as a whole think about the policy-making process based on interviews with just five people.

Types of data

The form that data (or information) takes can vary, from surveys to interview and focus group transcripts, ethnographic field notes, historical records and speeches, to government statistics, to name but just a few. Political scientists, perhaps more than researchers in any other discipline, draw on a wide and varied range of sources of information to study the political world. And just as the types of data used vary, the techniques used to analyse the data also vary. In this section we discuss the main types of data that are used in Politics and International Relations and provide a brief overview of the issues they raise with respect to validity, reliability, and selection bias.

It is common to distinguish between primary and secondary data. *Primary data* is often associated with data we collect ourselves. That is, data collected, recorded, and analysed by the researcher carrying out the project. By contrast, *secondary data* is often associated with data collected by other people. That is, the researcher analyses (or re-analyses) data that has been collected by someone else. The distinction between these two types of data relates more to the issue of who collected the data than it does to the type of data in question. For example, we could conduct a survey ourselves, in which case the survey would be an example of primary data. Or we could analyse a survey that has been conducted by someone else (such as the British Election Study), in which case the survey would be an example of secondary data. In both cases we are dealing with the same type of data, which has the same strengths and weaknesses.

The main advantage of collecting primary data yourself is that you have full control over data collection, and so are aware about any issues that might affect data quality (such as a low response rate to your survey). The main disadvantage is that data collection can sometimes be a time-consuming and costly exercise—and you may not have the resources to collect the data that you need. However, often we can use secondary data to answer the questions we are interested in. The UK Data Archive is an incredibly valuable resource, and contains records of thousands of studies (both quantitative and qualitative) on a wide variety of different topics that are freely and publicly available. The main disadvantage of using secondary data—or at least the main issue to be aware of—is that it is not always easy to tell whether the data is of high quality. It is therefore important to only ever use data you have confidence in, where there are clear methodological statements about how the data was collected, who it was collected by, and importantly, who it was funded by and why it was collected.

Data can be categorized in many different ways. If we want to find out information about something we have three main options. We can ask someone a question, we can observe what they do, or we can read what they have written. These different strategies form the main methods of primary data collection that we use in political science: interviews; observation; and document analysis—and we discuss each method in-depth in the following chapters. Within these broad categories of data collection there are many distinct approaches and we can distinguish between those that produce quantitative data and those that produce qualitative data (see Box 7.1). In addition, researchers frequently use *official statistics* or data that has been compiled by government agencies or international organizations (such as the World Bank). This type of secondary data raises a number of distinctive issues, which we discuss later.

BOX 7.1 Different Types of Data

Method of collection	Type of data	
	Quantitative	Qualitative
Interview	Results from a survey	Transcripts from interview or focus group
Observation	Roll call votes from parliament	Field notes from participant observation
Documents	Policy position of parties, salience of issues	Party manifestos, speeches, policy documents, diaries
Secondary sources	Government statistics	Government reports

Quantitative and qualitative data

One of the most common ways of classifying data is to distinguish between quantitative data, which is numerical, and qualitative data, which is not. Whereas quantitative researchers assign numerical values to their data, qualitative researchers tend to assign words (King, Keohane, and Verba 1994: 151). Thus, even when researchers are looking at the same source—for example, a party manifesto—they may choose to represent the data in different ways (see Chapter 14). The primary differences between qualitative and quantitative data thus relate to how the data are represented (or coded) rather than what type or source of data is necessarily being analysed.

Sometimes people get very animated about whether or not qualitative data is better than quantitative data. Some people might think that numbers tell you nothing, and so the only way you can ever find something out for sure is to do a detailed qualitative study and describe what you see using words. Others think that qualitative studies prove nothing, and so like to just rely on numbers. There are not many areas of life where we draw such firm distinctions between different types of evidence—and wilfully ignore what we don't like the sound of. Most of the time, when we care about getting the 'right' answer to a question we will consider a variety of different types of evidence—or data. Suppose we are buying a house, we might consider quantitative data (how much it costs; how big the house is) and qualitative data (what the house looks like, whether it is light and airy, or has a 'wow factor'). It would be a strange person that decided to buy a house without knowing how much it cost or without seeing what it looked like. Similarly, when we investigate a political phenomenon it makes sense to consider all types of relevant information regardless of whether it is qualitative or quantitative. Most examples of good research in Politics and International Relations draw on multiple sources of qualitative and quantitative data. And even when the data presented is explicitly quantitative, such as the results from a large-scale social survey, often a great deal of qualitative research has gone into the process of developing and testing the questions during various rounds of pilot studies.

That said, there are some important differences between qualitative and quantitative data, which involve a trade-off between internal validity and external validity. Qualitative

researchers try to achieve accurate measures, but they generally have somewhat less precision (King, Keohane, and Verba 1994: 151). Thus qualitative research is a very good way of assessing and establishing the validity of the data that we use—even if that data is quantitative in form. But since qualitative data tends to prioritize detail over theoretical generality—it is often very difficult to tell what the broader implications of the data are—and whether similar data on different cases would produce consistent results. Part of this is simply just a small-N problem (see Chapter 9). This can either be overcome gradually by **replication**, where similar data is collected in different contexts to see if the findings can be applied in the same way. Or it can be overcome by deriving testable hypotheses from the qualitative data that can then be examined more generally using quantitative data (see Chapter 13). In either case, qualitative and quantitative data can often be combined to good effect—and so the distinction between the two types of data is perhaps not as divisive as is often portrayed. Data also varies with respect to other characteristics—such as the **level of measurement**, and focusing purely on whether the data is qualitative or quantitative risks downplaying the salience of these other differences.

Levels of measurement

Quantitative analysis requires that the information or evidence that we have collected is converted into numbers. Sometimes it is fairly obvious what these numbers refer to, such as when we are recording a respondent's age or a country's gross domestic product (GDP) per capita. However, at other times it is not so intuitive, such as when we are recording a respondent's religion or a country's electoral system. How we interpret the numbers we assign to different variables then depends upon what the numbers actually refer to, and whether they are in a sense 'real numbers' or arbitrary codes for distinct or qualitative categories.

Broadly speaking, there are three levels of measurement: nominal (also called categorical), ordinal, and interval (also called continuous or scale). We'll now describe each of these.

'Nominal' comes from the Latin for name. A **nominal** variable is one where the numbers assigned to the variable are interesting only in so far as the labels—or names—that are attached to them are interesting. To interpret the variable, we must know what the values refer to and the names of the different categories. For example, consider the following questions from the British Election Study. We can see that each response to the question is assigned a number, but these numbers do not refer to anything other than the label that they represent.

> Talking to people about the general election on June 8th, we have found that a lot of people didn't manage to vote. How about you? Did you manage to vote in the general election?
>
> 1. Yes, voted.
> 2. No.
>
> [IF YES] Which party did you vote for in the general election?
> [DO NOT PROMPT]
>
> 1. Conservative
> 2. Labour
> 3. Liberal Democrat

4. Scottish National Party (SNP)
5. Plaid Cymru
6. Green Party
7. United Kingdom Independence Party (UKIP)
8. Other party.

The fact that people who didn't vote are coded 2 and people that did vote are coded 1 does not mean that non-voters are twice as apathetic as voters. Although each of these responses is assigned a numerical value, the number is of no interest itself, we are only interested in what the number refers to. Nominal variables are therefore often considered to contain a more qualitative dimension, as they refer to distinct categories that matter in quality or attribute rather than in quantity or magnitude (Lewis-Beck, Bryman, and Liao 2011).

By contrast, in the case of **ordinal** variables, the numbers assigned to the different response categories do have some meaning. They have an order. Consider the following example:

Let's talk for a few minutes about politics in general. How much interest do you generally have in what is going on in politics?

1. None at all
2. Not very much
3. Some
4. Quite a lot
5. A great deal.

Once again, each response category is assigned a value, but this time we can interpret the values according to the order in which they are arranged. We can think of the numbers as referring to levels of political interest, so someone who replies that they have not very much interest in politics is coded 2, and from this we can see that they have less interest in politics than someone whose response is coded 4. But there is a limit to the extent that we can draw meaningful conclusions about how much more or less interested they are. We cannot say that someone who is coded 4 is twice as interested in politics as someone who is coded 2. Nor can we say that the difference in interest between 4 and 5, which is 1, is the same as the difference in interest between 1 and 2, which is also 1. In short, we can only make meaningful statements about the order of the responses, not the magnitude of the differences between them.

With **interval**—or scale—variables, the numbers do make sense. Indeed, there is no distinction between the value and the label. An example of this type of variable is age.

Now, a few questions about yourself and your background. What was your age last birthday?

Age in years __

If someone is 34 years old, they are coded as 34. The numbers have an order, so someone who is 40 is older than someone who is 30, and the numbers also have a magnitude that we can interpret, so someone who is 60 is twice as old as someone who is 30. The difference

between 25 and 30 is five years, and this distance is the same as the difference between someone who is 45 and 50. For this reason, interval variables are often referred to as real numbers.

Interview data

Interview data is perhaps one of the most widely used forms of data in political science. Much of our research concerns what people (or actors) do, say, or think. One particularly useful way of finding out this information is to ask questions to those concerned. There are many different ways to do this. Interview data can be gathered using more, or less, structured methods: for instance, using open-ended and flexible questions in semi-structured interviews or focus groups (Chapter 12), or questions to be answered by respondents selecting from among a fixed set of choices in surveys (Chapter11). The main instrument for gathering data might be the researcher himself, or it may be that data are not 'collected' at all, but rather are co-produced by the researcher and the subject as part of an ethnography (Chapter 13). Interviews might be carried out with a representative cross-section of a given population, or with experts who have specialized knowledge about a given topic.

The actual data that these methods of collection produce are the specific responses to the questions that respondents provide. These responses might be recorded in different ways: they could be audio recordings on digital devices; they may be transcribed verbatim in written documents; or they may be coded using numbers in databases. For example, a researcher may be interested in the opinions of young people towards electoral reform. He could try and find out about these opinions by collecting data through surveys, or through interviews, or focus groups, or by ethnography. In each case the researcher could ask the same question. For example: 'what do you think about the House of Lords? Do you think it needs reform?' This is obviously a very general open-ended question which people might interpret in a number of ways and give a variety of different answers to. Some people might say that they don't know or that they've never thought about it, or they have no idea. We could write down each of these responses word for word. We might think that this sort of data is qualitative; but if we decide to code each of the answers into broad response categories, such as 'no opinion', we can easily ascribe a numerical value to each category and transform the qualitative data into quantitative data. Different types of interview data can therefore be integrated to give a fuller picture of what young people actually think.

Observational data

Whereas interview data is based on the ancient art of asking questions to find out what people think, say, and do, observational data is based on something rather different. It recognizes that what people say they do, and what they actually do are often different. For example, someone may claim that they are not sexist, but nonetheless may frequently behave in a sexist way. Accordingly, to get a 'true' sense of what people think and say and do, it is not enough to merely ask people questions and record their answers; it is also necessary to observe what people do in practice.

There are various ways in which political scientists do this. One form of observational data comes from participant observation (Chapter 13). For example, in a classic political

ethnography Richard Fenno (1978) conducted research in the USA with eighteen members of Congress as they travelled around their constituencies, spending time with them and observing what they did and who they spoke to on a day-to-day basis, providing fascinating insights in to how Congress members interact with their constituents, and how this translates into their legislative effectiveness. Another form of observational data comes from collecting systematic data on the behaviour or actions of different actors. This data may be collected by researchers or recorded by different organizations. For example, Keith Poole and Howard Rosenthal (2000) analysed sixteen million individual roll call votes in the Houses of Congress since 1789 to examine the voting patterns of legislators in the USA. Another form of observational data comes from experiments—when participants are exposed to different stimuli to see how they react (see Chapter 8). Different types of observational data can therefore be integrated to give a fuller picture of how people actually behave.

Documents

With interviews and observation the data is collected—or generated—by the researcher. That is, through asking questions and recording answers, or by observing and recording patterns of behaviour, the researcher is able to produce new data that can be analysed in order to answer the key research questions. Other types of data do not need to be collected in the same way. For example, researchers may analyse written policy documents or political speeches, media reports or broadcasts, personal diaries—or, social media posts from both a qualitative and a quantitative perspective (see Chapter 14). Analysing documents is generally an unobtrusive method of data collection (though this is not quite so clear-cut when it comes to social media posts). In this case the raw data is the document itself. Sometimes researchers will analyse the documents from a qualitative perspective, and extract excerpts or quotations to illustrate a point. Other times researchers will analyse the documents from a quantitative perspective and code the content of the document into distinct categories to produce new data.

For example, the Manifesto Project provides the research community with parties' policy positions derived from a content analysis of parties' electoral manifestos (Volkens et al. 2015). It covers over 1,000 parties from 1945 until today in over fifty countries on five continents. In each manifesto, the policy statements in each ('quasi-') sentence are classified into fifty-six policy categories over seven policy domains. The coded data provide a useful indication of party positions since they represent the choices that the electorate faces before each election. These measures are generally consistent with those from other party positioning studies, such as those based on expert placements, citizen perceptions of parties' positions, and parliamentary voting analyses, which provides additional confidence in the validity and reliability of these estimates (Laver, Benoit, and Garry 2003).

Secondary data

Whereas the types of data discussed earlier can be primary or secondary, other types of data are purely secondary. Political scientists frequently use data or sources that they have not collected themselves, and which have been collected or produced for purposes other than academic research. This type of data often relies upon the record-keeping activities

of government agencies, interest groups, and international organizations (see Johnson and Reynolds 2008: chapter 9). For example, various government agencies routinely publish official statistics which document the level of poverty, unemployment, and crime within the country. Although these government statistics can be useful, they also need to be treated with caution. As with all secondary data, the researcher is not directly involved in the process of data collection, and so is in a sense at the mercy of the practices and procedures that have been adopted by the organization that has collected the data. These practices and procedures are not always transparent. One particular issue to be aware of is bias. Government statistics on issues such as crime or poverty can be very sensitive—and in some cases government agencies may try to 'massage' the figures to present the government in a more favourable light. This can make comparative research very difficult—as governments in some countries may simply be more willing to publish accurate data than others. Moreover, changes in the way crime or poverty are classified can affect the level which is recorded. This can make analysis difficult over time. As always then, when using secondary data it is important to assess its validity and reliability and to be aware about potential biases.

Big data

So far we have considered the main types of data that are used in Politics and International Relations. In recent years however, there has been a much heralded 'revolution' in the way that data is collected and analysed. This so-called 'big data' revolution challenges many of our assumptions about what data is, how data is generated, and what data can tell us about the social and political world. In this section we discuss what big data is, how it is being used, and its potential applications for research in Politics and International Relations. We also examine some of the problems and pitfalls with big data, and discuss the differences between big data and other types of data.

Big data is a very modern phenomenon. Although political scientists have long worked with very large datasets—such as Census data which contains data on millions of people—it is only relatively recently that people have started to speak about 'big data' as a separate category of data. As Burt Monroes (2012) explains, as political scientists we are used to thinking about 'small-N' or qualitative studies and 'large-N' quantitative studies, and so it is perhaps natural to equate big data with 'really large-N' studies and to think about it purely in terms of scale and size. However, this is to misunderstand the 'big' in big data and the qualitative difference that exists between 'big' data and other types of data. At the turn of the century, IT analyst Doug Laney laid out a framework for understanding the challenges that were then emerging in data management for business, now widely referred to as 'The Three Vs of Big Data': Volume, Velocity, and Variety (Laney 2001). Simply put, data were beginning to be created at a scale, speed, and diversity of forms not seen before. According to Gary King the 'big data revolution' is a technological revolution that has transformed our ability to produce, collect, store, and analyse vast amounts of data. One of the key features of big data is that the data is generated almost spontaneously. Vast amounts of activity now leave a digital footprint, and these digital records constitute a sort of data that can be readily harvested. Researchers do not therefore need to generate or collect the data in the traditional sense. For this reason big data is sometimes described as 'found' data. The data

BOX 7.2 Uses of Big Data (King 2013)

Topic	Traditional data	Big data
Opinions of political activists	Survey—how interested are you in politics?	Analysis of Twitter: billions of political opinions in social media posts
Exercise	Survey—how many times did you exercise last week?	500,000 people carrying mobile phones with accelerometers
Social contacts	Survey: 'Please tell me your five best friends'	Continuous record of phone calls, emails, text messages, social media connections
Economic development in developing countries	Dubious government statistics	Satellite images of human-generated light at night, road networks, other infrastructure

Source: Reproduced with permission from Gary King. 'Big Data is not about the Data!'. *Talk at Harvard University Information Technology Town Hall*, 28th October 2013. http://gking.harvard.edu/files/gking/files/evbase-huit.pdf?m=1382033115.

is already there, and all the researcher needs to do is harvest it. This also makes big data incredibly cheap compared to similar more traditional forms of data. Traditional face-to-face surveys are very expensive. The British Election Study for the 2015 elections cost over £1 million. By contrast, the opinions of millions of people on Twitter are free to harvest.

As Gary King (2013) has said, big data offers the opportunity to study things we have always looked at from a new perspective and to study things we previously found difficult to analyse (see Box 7.2). Perhaps one of the most widespread uses of big data is to mine social media. Traditionally, if we were interested in the political opinions of people we would carry out a survey—of say 1,000 people—and ask them what they thought about a particular issue. This form of data is not only relatively expensive to collect, but also time-consuming. Now, with social media, millions of people readily volunteer their opinions on matters of the day. These opinions can be freely harvested from platforms such as Twitter in a matter of seconds. Scholars such as Ben O'Loughlin have used this approach to investigate how people on Twitter respond in real-time to high-profile political events. One such event he looked at was the controversial appearance of Nick Griffin, the former leader of the far-right British National Party, on the weekly British political debate show BBC Question Time (O'Loughlin and Anstead 2011). During the course of the TV show, some 50,000 tweets were posted about the programme which were then harvested and coded, revealing which segments of the programme produced the biggest 'spikes' in Twitter activity.

Other examples of big data involve harnessing mobile technology. Traditionally, if we wanted to know how much exercise people took the only way we could find out was by carrying out a survey and asking people what forms of activity they undertook. Now, researchers are using mobile phones with special apps to actually record how much hundreds of thousands of people move around each day, measuring the distance they travel and the speed at which they move using accelerometers. Indeed, there are many innovative ways in which big data is being used and the digitalization of vast catalogues of archival materials is blurring the boundary between qualitative and quantitative research and forging new collaborations between historians and computer scientists. For example, one ingenious study was carried out by Dan Cohen and colleagues on trial records from the Old Bailey from the

seventeenth century as part of the Digging into Data programme. The Old Bailey website (http://www.oldbaileyonline.org) makes available 127 million words of accurately transcribed historical text, recording the details of 197,000 trials held at the Old Bailey or Central Criminal Court in London, between 1674 and 1913 (Hitchcock and Shoemaker 2006). All manner of people, across all ages, genders, and classes appeared in London's court. Their words, collected in the evidentiary statements (comprising 85 per cent of the proceedings), detail crimes, their ancillary activities, their perpetrators, and their victims. Dan Cohen, for instance, looked at cases that involved women in the nineteenth century and was able to chart changing attitudes towards gender by examining how women were punished for different offences. By looking at cases at the Old Bailey involving bigamy he could see a significant rise in women taking on other spouses when their husbands had left them. It also became very clear that these late-Victorian women were not receiving significant punishment for this action, as they had earlier in the century.

Although big data opens up many new avenues for research, it also presents some familiar problems to do with reliability, validity, and bias. Just because there is more data, it does not mean that these problems go away, and indeed, they can even be amplified. Perhaps one of the biggest problems with big data is to do with selection effects. Big data is often seen in a favourable light by policy-makers because it provides a low-cost way of gathering information which can then be used to allocate resources. However, failing to take selection issues into account can lead to serious problems, and inequitable outcomes. A good example of this comes from Boston in the USA which had a big problem with potholes in the city streets, with 20,000 potholes needing to be filled every year. This was a considerable drain on the city finances, not least in terms of identifying where the potholes were in the first place. To help allocate its resources efficiently, the City of Boston released the Street Bump Smartphone app, which draws on accelerometer and GPS data to help detect potholes, instantly reporting them to the city. This was an ingenious use of big data technology. However, there was a problem. People in lower income groups in the US are less likely to have Smartphones, so the people using the app tended to be well off, and consequently the potholes identified tended to be in well-off areas or on popular commuter routes. This meant that the potholes in poorer parts of the town were less likely to be noticed and therefore less likely to be fixed—hardly an equitable approach. Fortunately this problem was soon identified and so measures were put in place to try and redress the digital divide.

Other examples of big data failing to live up to its initial promise are illustrated by the now famous parable of Google Flu trends (Lazer et al. 2014). Outbreaks of flu—or mini epidemics—are common in many countries, and these can put a strain on medical services. Identifying when and where outbreaks are likely to occur can therefore be an enormous help for distributing resources to where they are most needed. Traditionally local doctors would report cases as they came in—but this process tended to be slow and so did not provide much of an early warning signal. Google realized that outbreaks of flu could be predicted with a remarkable degree of accuracy by looking at the search terms that people entered into Google. But the problem was that this was a largely inductive approach. Google identified the search terms that best predicted outbreaks of flu, whether those terms were directly related to the symptoms of flu or not. There was thus little theory driving the data collection. Although the algorithm they developed was highly successful at first, it soon stopped working, serving as a timely reminder that correlation does not necessarily equate to causation.

The approach that we have described in this book is theory-led. We come up with an idea—we engage with theory to form hypotheses—which we then try and operationalize in order

to test. Some commentators have suggested that big data can replace this scientific method, with theory and research design essentially becoming obsolete. The editor-in-chief of *Wired Magazine*, for example, writes that 'Correlation is enough. …We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot' (Anderson 2008). However, this approach is almost certainly doomed to failure over the long run. Political scientists have typically had more reservations about the promise of big data than their computer scientist counterparts, largely because of their training in statistical theory, which makes them aware of issues related to things like sampling populations, confounders, overfitting, and multiple hypothesis testing (Clark and Golder 2015). As Manski (2013) and Keele (2015) observe, conclusions result from a combination of data and theoretical assumptions—without assumptions, data is just data. As Clark and Golder (2015) put it, the choice confronting the researcher is therefore not between induction or deduction, or description and explanation, but between observation guided by conscious theorizing and observation guided by unconscious theorizing.

Data quality checklist

Whatever type of data is collected or used there are a number of fundamental principles worth addressing to convince yourself and your reader that the data is of good quality and that the findings can therefore be trusted (King, Keohane, and Verba 1994: 23). The process of establishing the validity and reliability of our data is important for determining whether or not we can have confidence in our results. Because we use the data to test our hypotheses we must have confidence that the data is as accurate as possible. If our data is not accurate then we may come to faulty conclusions, and either incorrectly reject a hypothesis when it is in fact true, or incorrectly accept a hypothesis when it is in fact false. Obviously we want to have confidence in our own research, but we also want others to have confidence in it too and to take the findings seriously.

The first and most important guideline is to be transparent about how the data have been collected and to report and record in as much detail as possible the processes by which the data have been generated. One way in which we can give others confidence in what we have done is to be as open and transparent as possible about what we have done at each stage of the data collection and analysis process. In this respect, the principles of replication and verification are very important. It is now a widely accepted practice to make all publicly-funded data publicly available. This is now a funding requirement in many countries around the world. Public bodies fund research and so demand that the data be shared amongst the community. This means that the data collected by individual researchers or teams of researchers is freely available to other scholars who may wish to analyse it. Although quantitative research has led the way in making data publicly available, qualitative researchers are now also being encouraged to be more transparent about their data collection. Datasets and interview or focus group transcripts with detailed methodological appendices are now routinely uploaded to project websites and are deposited in digital archives such as the UK Data Archive.

Making data available in this way creates a public resource that other scholars can share. It also makes the research process transparent. **Verification** refers to the practice of making data available for secondary analysis so other scholars can re-analyse the data to see if they

come to the same conclusions. Replication refers to the practice of carrying out a similar data collection exercise to see if the findings from the original study can be reproduced on an independent data set. Replication—whether it is for qualitative or quantitative data—is the primary way in which we can establish the external validity of our data.

These principles of replication and verification make the research process more transparent, which in turn gives us confidence in the results as it adds an extra layer of scrutiny to how the data has been collected and analysed. This openness and transparency allows us to develop knowledge by building on each other's work. It also allows us to identify cases of misconduct. This scrutiny has helped to uncover some major instances of research fraud. In a now famous case, in 2014, an article was published in *Science* by PhD student Michael LaCour and Professor of Political Science Donald Green, which claimed to show that personal contact with gay rights activists had long-lasting effects on people's attitudes towards gay marriage. In the process of trying to replicate the study—and re-analyse the data that had allegedly been collected, another PhD student, David Broockman, noticed some irregularities which led him to suspect that the data had been fabricated. These suspicions were corroborated when the survey agency who had apparently collected the data were contacted said that they knew nothing about the study, and that they had never worked with Michael LaCour (who was the principal investigator in the study). The article was later retracted from the journal. This scandal undoubtedly did a lot of damage to the credibility of the discipline. However, the fact that it was uncovered shows the value of replication and verification. Although replication and verification cannot ensure that irregularities do not take place, they can help to uncover them. And without replication and verification it is highly likely that more fraud would occur. As in other aspects of political life, transparency is the best antidote to corruption.

Conclusion

Politics and International Relations studies are about making arguments and supporting those arguments with evidence. The data that we collect provides the information base for these arguments. If the evidence is bad then the argument is misleading. In order to have confidence in our results it is important to ensure that our data is as good as possible. And so that others may also have confidence in what we have done, it is important to be as transparent and open as possible about how our data have been collected and how they have been coded. Although there is a tremendous amount of different types of data that we can draw on, the principles of what makes good data are similar. Whether we are interested in qualitative interviews or big data analytics, the issues of selection, measurement, theory, and validity are equally important.

Questions

1. How can we measure difficult-to-define concepts in Politics and International Relations?
2. Why is measurement important?
3. Is there a trade-off between internal validity and external validity? Which should we be most concerned about?

4. What is the difference between qualitative and quantitative data? How important are these differences?
5. Is it always better to collect data on a large number of cases? When might data from just a few cases lead to valid conclusions?
6. How would you assess the validity and reliability of secondary data? What questions would you want to ask?
7. Is big data different to other types of data? What, if anything, is distinctive about it?
8. Should data be made publicly available? What are the potential drawbacks to making data available?

Guide to Further Reading

Clark, W. and M. Golder (2015), 'Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science?', *PS: Political Science and Politics* 48(1): 65–70.
This special issue contains a number of articles which consider some of the problems and pitfalls of big data and the challenges it presents to conventional methodological approaches in political science.

Coppedge, M. (2012), *Democratization and Research Methods* (New York: Cambridge University Press).
Provides an in-depth but accessible discussion about the problems of operationalizing and measuring democracy, and the consequences for analysis.

de Vaus, D. (2002), *Surveys in Social Research* (London: Routledge).
Although this book is primarily about surveys, it contains very clear discussion about measurement issues, operationalization, and the principles of validity and reliability, which are of relevance to other methods of data collection.

Heath, A., L. Richards, and N. D. de Graaf (2016), 'The Sociology of Corruption', *Annual Review of Sociology* 42.
Provides a detailed discussion about the strengths and weaknesses of different approaches to measuring corruption, and outlines promising avenues for further research.

King, G., R. Keohane, and S. Verba (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ: Princeton University Press).
This classic text provides a thorough examination of measurement and selection issues in qualitative and quantitative research.

References

Anderson, C. (2008). 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', *Wired Magazine*, 16 July.

Binder, M. (2015), 'Paths to Intervention: What Explains the UN's Selective Response to Humanitarian Crises?', *Journal of Peace Research* 52(6): 712–26.

Clark, W. and M. Golder (2015), 'Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science?', *PS: Political Science and Politics* 48(1): 65–70.

Coppedge, M. (2012), *Democratization and Research Methods* (New York. Cambridge University Press).

Dahl, R. (1989), *Democracy and Its Critics* (New Haven, CT: Yale University Press).

de Vaus, D. (2002), *Surveys in Social Research* (London: Routledge).

Fenno, R. (1978), *Home Style: House Members in Their Districts* (New York: Harper Collins).

Glaeser, E. L. and R. E. Saks (2006), 'Corruption in America', *Journal of Public Economics* 90(6): 1053–72.

Goel, R. and M. Nelson (2011), 'Measures of Corruption and Determinants of US Corruption', *Economics of Governance* 12(2): 155–76.

Heath, A., L. Richards, and N. D. de Graaf (2016), 'The Sociology of Corruption', *Annual Review of Sociology* 42.
Hitchcock, T. and R. Shoemaker (2006), 'Digitising History from Below: The Old Bailey Proceedings, 1674–1834', *History Compass* 4(2): 193–202.
Johnson, J. and H. T. Reynolds (2008), *Political Science Research Methods* (Washington, DC: CQ Press).
Keele, L. (2015), 'The Discipline of Identification', *PS: Political Science and Politics* 48(1): 102–6.
King, G. (2013), 'Big Data Is Not about the Data!' Talk at Harvard University Information Technology Town Hall, 28th October.
King, G., R. Keohane, and S. Verba (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ: Princeton University Press).
Laney, D. (2001), '3D Management: Controlling Data Volume, Velocity, and Variety', *Application Delivery Strategies* (META Group, Inc.), 6 February.
Laver, M., K. Benoit, and J. Garry (2003), Extracting Policy Positions from Political Texts using Words as Data', *American Political Science Review* 97: 311–31.
Lazer, D., R. Kennedy, G. King, and A. Vespignani (2014), 'The Parable of Google Flu: Traps in Big Data Analysis', *Science* 343(6176): 1203–5.
Lewis-Beck, Michael S., Alan Bryman, and Tim Futing Liao (2011). *The SAGE Encyclopedia of Social Science Research Methods* (Sage: Thousand Oaks).
Manski, C. F. (2013), *Public Policy in an Uncertain World: Analysis and Decisions* (Cambridge, MA: Harvard University Press).
Monroes, B. (2012), 'The Five Vs of Big Data Political Science. Introduction to the Virtual Issue on Big Data in Political Science', *Political Analysis* 19: 66–86.
Moser, Sir C. and G. Kalton (1971), *Survey Methods in Social Investigation* (London: Heinemann Educational Books).
O'Loughlin, B. and N. Anstead (2011), 'The Emerging Viewertariat and BBC Question Time: Television Debate and Real Time Commenting Online', *International Journal of Press/Politics* 16(4): 440–62.
Poole, K. and H. Rosenthal (2000), *Congress: A Political Economic History of Roll Call Voting* (Oxford: Oxford University Press).
Vanhanen, T. (1997), *Prospects of Democracy: A Study of 172 Countries* (London: Routledge).
Vanhanen, T. (2000), 'A New Dataset for Measuring Democracy, 1810–1998', *Journal of Peace Research* 37(2): 251–65.
Volkens, A., P. Lehmann, T. Matthieß, N. Merz, S. Regel, and A. Werner (2015), The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). Version 2015a (Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB)).

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

Part 3

How to do Research in Practice

8	**Experimental Research**	211
9	**Comparative Research**	231
10	**Historical Research**	260
11	**Surveys**	284
12	**Interviewing and Focus Groups**	309
13	**Ethnography and Participant Observation**	339
14	**Textual Analysis**	364
15	**Quantitative Analysis: Description and Inference**	393
16	**Patterns of Association: Bivariate Analysis**	422
17	**A Guide to Multivariate Analysis**	451

Experimental Research

Chapter Summary

This chapter examines the principles of experimental research design and discusses the issues and problems associated with different aspects of the approach. In doing so we pay special attention to the issue of internal and external validity, the common obstacles associated with experimental research, and what can be done to try and avoid or minimize them. The chapter examines:

- basic principles of experimental research;
- field experiments;
- laboratory experiments
- natural experiments;
- internal and external validity;
- problems;
- ethical issues.

Introduction

The use of experimental designs to study political phenomena has grown considerably in recent decades. Although there is a long history of experimental political research, with some early pioneering studies dating back to the 1920s and 1930s (see in particular Harold Gosnell 1927), in recent years there has been a marked increase in the use of this approach. Indeed, experimental research is now perhaps the fastest growing area of political research. Researchers in politics have used experimental designs to study political mobilization (Gerber and Green 2000; John and Brannan 2008; Townsley and Turnbull-Dugarte 2019), voting (Lodge, McGraw, and Stroh 1989), election fraud (Mack and Stoetzer 2019), negative campaigning (Wattenberg and Brians 1999), coalition bargaining (Fréchette, Kagel, and Morelli 2005), electoral systems (Morton and Williams 1999; Song 2019), clientelism (Wantchekon 2003; Heath and Tillin 2018), culture (Henrich et al. 2004), identity (e.g. Habyarimana et al. 2007), foreign policy decision-making (Geva and Mintz 1997), international negotiations (Druckman 1994), justice (Frohlich and Oppenheimer 1992), and deliberation (Simon and Sulkin 2001).

The appeal of the approach is easy to appreciate. Experimental designs are often thought to most closely resemble the true scientific method (see Chapter 2), and as such

they are widely regarded as being the most effective design for testing whether or not two variables are causally related. They manage to do this thanks to the rigorous use of experimental control. This helps researchers to isolate the impact of a specific variable and to overcome one of the main problems that researchers face when they want to investigate causal relationships. Since there is a vast number of potentially important variables that may influence any given political phenomenon, in the absence of rigorous controlled experiments (not dissimilar to those used in medical research) it is very difficult to know which variable is responsible for causing the phenomena under investigation. One of the great strengths of experimental research is that the researcher can control the environment and manipulate particular variables of causal interest with great precision. One of the defining characteristics of experimental research is **intervention** by the researcher in the data-gathering process. Morton and Williams (2008) describe this as 'playing God'. We call the data gathered by intervention (or **treatment**, as it is sometimes known) experimental data.

In this chapter we discuss some of the main issues in experimental research and some of the main ways in which the approach has been used in the study of politics. In the first part of this chapter we provide a brief overview of the basic principles of experimental design, and how experiments can be used to test causal hypotheses. We then go on to discuss some of the main ways in which experimental research works in practice. Experiments can be carried out in a number of different locations. Experiments can be carried out in a laboratory, in the field, and by utilizing natural occurrences. In laboratory experiments, the subjects (the participants in the study) are recruited to a common location where the intervention takes place. In field experiments, the intervention by the researcher takes place in real-world environments, and subjects may not even know that they are participating in an experiment. This obviously raises some ethical issues (which we discuss in detail in Chapter 6). By contrast, natural experiments rely on naturally occurring events as interventions rather than interventions controlled by the researcher.

We then discuss some of the main issues to do with the reliability and validity of experimental research. This has to do with whether or not we are confident that a causal relationship really exists between two variables (sometimes known as internal validity) and whether or not we are confident that our findings can be generalized to subjects outside the confines of the experiment (sometimes known as external validity). We will also discuss some of the practical and ethical problems associated with doing different types of experimental research.

Basic principles of experimental design

Experimental research relies on the use of control groups (in which no intervention takes place) and experimental groups (in which interventions do take place) and the **random assignment** of subjects (participants) to control and experimental groups.

The control group provides a point of reference to which the effect of the intervention can be compared. The random assignment of subjects to control and experimental groups ensures (as far as possible) that the two groups are similar to each other (see Chapter 11 on

sampling). These groups are then treated in the same way in every respect apart from the intervention that is carried out on the experimental group. Any differences that are then observed between the groups on the outcome variable of interest (the dependent variable) can then be attributed to the intervention that took place.

The classic version of the experimental design comprises five steps (see de Vaus 2001: 48):

1. Two groups: one group that is exposed to the intervention (the experimental group) and one group that is not exposed to the intervention (the control group).
2. Random allocation of subjects to the groups before the pre-test.
3. One pre-intervention (pre-test) measure on the outcome variable.
4. One intervention (test/treatment).
5. One post-intervention (post-test) measure on the outcome variable.

Table 8.1 provides a conceptual overview of this method. The columns refer to the groups to which subjects are randomly assigned. Once subjects have been assigned to a control group and an experimental group, both groups are pre-tested on the outcome variable of interest, the dependent variable (Y). This provides a baseline measure which later results can then be compared to. In the next step, a treatment (or an intervention) is administered to the experimental group but not the control group. The treatment refers to the key independent variable (X) of interest. This is related to the causal hypothesis which we wish to test. To see whether the hypothesis is confirmed or not we need to carry out a post-test on both groups on the dependent variable. If our hypothesis is supported, then we should observe that the test statistic for our dependent variable Y has changed for the experimental group (which received the treatment) but has not changed for the control group. The test for the effect of the intervention is therefore carried out by comparing changes in the experimental group before and after the intervention to changes (if any) in the control group.

This classic design is often simplified in practice by dropping the pre-test stage. Valid inferences about the causal significance of the treatment variable can still be drawn so long as the groups are large enough. The key thing is that subjects are randomly allocated to the experimental and control groups. In effect this means that any differences between the experimental and control groups are random and will not account for group differences in outcomes (see de Vaus 2001: 60). With this sort of design, rather than looking at the *amount of change* between the groups, the analysis is based on the post-test *differences* between the groups. Because subjects have been randomly allocated to their groups, the post-test differences should be the same as the difference in change scores of the experimental and control groups in the classic set-up.

Table 8.1 Experimental design

	Control group	Experimental group
Pre-test	Measure on outcome variable (Y)	Measure on outcome variable (Y)
Intervention	No treatment	Treatment
Post-test	Measure on outcome variable (Y)	Measure on outcome variable (Y)

There are also a number of issues to do with **case selection** in experimental research. The first issue relates to how subjects (or participants) are selected for the study. The second issue is to do with how subjects are assigned—or allocated—to control or experimental groups in the study. Generally speaking, researchers who carry out experiments are more concerned with assignment than selection, particularly in laboratory experiments.

It is the random assignment of subjects to groups, rather than the random selection of subjects in the first place, that is most important for testing causal hypotheses. For this reason, many laboratory experiments do not rely upon representative or random samples of subjects. It can be very difficult, time-consuming, and expensive to select a random sample of participants from across the country and then transfer them to the location of the laboratory. It is often far more convenient to recruit people to the study who live near the lab. Indeed, since laboratory experiments frequently take place on university campuses where academics are based, the most convenient and accessible sample is often an all-student sample. This is precisely what a lot of experimental research relies upon. As long as the students are randomly assigned to treatment and control groups, the fact that they are not representative of the wider population does not undermine the internal validity of the study (see Druckman and Kam 2011 for an extended discussion on this topic). However, it might compromise the external validity of the study and the extent to which the findings can be generalized, at least in so far as the findings can be generalized to non-students.

We will now illustrate how these designs are carried out in practice with reference to some specific examples. In particular, we focus on laboratory, field, and natural experiments. The main purpose of this chapter is to provide a brief overview of the three main types of experimental design. Politics researchers have conducted a great deal of both laboratory research (e.g. on the impact of campaign commercials) and field experiments (e.g. on the effects of canvassing, telephone calls, and direct mail on voter turnout: Gerber and Green 2000; John and Brannan 2008). And, where possible, they have also exploited 'natural experiments' (e.g. on the effects of gender quotas on female political representation). In the following sections, we discuss each of these experimental design types in turn.

Laboratory experiments

In laboratory experiments, the subjects are recruited to a common location where the experiment takes place. The laboratory experiment is designed to ensure that the researcher has as much control as possible over the environment to which subjects are exposed. In this way, the experimental group and the control group can be exposed to exactly the same environment except for the experimental intervention. Accordingly, any difference that is recorded on the post-test measurement of the dependent variable can be confidently attributed to the presence (or absence) of the intervention.

The controlled environment of a laboratory is particularly well-suited to examining a wide range of political phenomena that might otherwise be difficult to investigate. In particular, Morton and Williams (2008: 346) suggest that laboratories have three main strengths as a setting for experimental research. Laboratory experiments allow the researcher to have a great deal of control over what the subject is exposed to. In a natural setting, people might be exposed to many different types of stimuli each day, so it is difficult to investigate what

impact, if any, each specific stimulus has on their attitudes or behaviour. In a laboratory, different stimuli can be manipulated one at a time, holding everything else constant. This allows causal hypotheses to be tested with far more precision. Second, laboratory experiments also allow the researcher to have a great deal of control over what variables are manipulated, and even to manipulate variables that might be difficult to vary in the real world. Third, laboratory experiments allow the researcher to create environments that simply don't exist in the real world. For example, they can explore how decision-making is influenced by entirely new voting systems.

With the rise of internet-based survey experiments, the distinction between laboratory experiments and survey experiments is becoming more blurred (see Morton and Williams 2009), and in some respects online survey experiments can be regarded as a type of laboratory experiment. The main difference between them is that in laboratory experiments subjects are recruited to a common location to complete the survey experiment, whereas in online experiments subjects are free to complete the survey experiment from any location they choose and so can reach a much broader audience. Surveys are frequently used to carry out question-wording experiments, where some respondents are asked a question in a certain way and other respondents are asked the same question but using slightly different words. These types of experiments can be used to examine how respondents react to different types of information, and information framed in different ways. We discuss some of these types of experiments in our chapter on surveys in Chapter 11, but we also provide a brief example here.

Thanks to the availability of free-to-use online survey platforms, such as PsyToolkit, which is run by the University of Glasgow, online survey experiments are a cost-effective way for students and other researchers to carry out original research. One such student-led survey experiment has been conducted on public attitudes towards refugees from Syria. The civil war in Syria provoked a humanitarian crisis which led to the displacement of millions of civilians. Many of these civilians embarked on a treacherous journey to seek sanctuary in Europe, and often ended up in refugee camps, such as the 'Jungle' in Calais, which became the temporary home for thousands of people hoping to get into the UK. The political response to this crisis was a source of great controversy. In a speech in the House of Commons, the then UK Prime Minister David Cameron referred to these people in the camps at Calais as a 'bunch of migrants'. This choice of language was widely criticized, and a spokesperson for the Labour Leader said the Prime Minister's comments were evidence of a 'wholly contemptible' attitude towards refugees. This was not the first time that Cameron had been accused of using inflammatory language on the subject. Earlier he had expressed concern about 'swarms of migrants' from the Middle East trying to 'break into Britain' from Calais. This led to accusations by the then Deputy Leader of the Labour party, Harriet Harman, that his choice of language was 'dehumanizing'.

The controversy that these episodes provoked inspired our students in the Department of Politics and International Relations at Royal Holloway, University of London, to examine the impact of language on public attitudes. In particular they examined whether referring to people as 'migrants' or 'refugees' affects how much support people have for letting Syrians come and live in the UK. The reasons why word choice might make a difference have a certain intuitive appeal. The term 'migrant' is frequently associated with people coming to the UK to look for work, and it has developed a number of negative connotations. By contrast,

the term 'refugee' refers to people in genuine need who are fleeing from persecution. All other things being equal, we might therefore expect people to be more sympathetic to the plight of refugees than that of migrants, even when the words are being used to describe exactly the same group of people.

In order to test this hypothesis the students carried out an online survey experiment with PsyToolkit. The sample was recruited by the students using Facebook, Twitter, and Group Chat. Although the sample itself was not randomly selected, crucially, for the purpose of the experiment, respondents were randomly assigned to one of the two treatment conditions. The first condition involved asking the following survey question: 'Do you agree or disagree that Britain should allow more *migrants* from Syria to come and live in the UK?' The second condition asked an identical question, but substituted the word 'refugee' for 'migrant': 'Do you agree or disagree that Britain should allow more *refugees* from Syria to come and live in the UK?' In both cases respondents could answer: strongly disagree, disagree, agree, or strongly agree. Over 1,850 people completed the survey.

Table 8.2 shows the results. Although the differences between the two groups are not great, they are statistically significant (see Chapter 15). Respondents were more likely to 'strongly agree' with the statement that refugees from Syria should be allowed into the UK than with the corresponding statement about migrants (23 per cent vs. 17 per cent). The difference between the two groups of six percentage points is not trivial. Moreover the mean level of agreement between the two groups is statistically significant. Talking about 'migrants' instead of 'refugees' therefore appears to have a negative influence on public opinion.

However, these types of experiments are not without their problems. An example of a more traditional laboratory experiment can help to illustrate both the potential strengths and weaknesses of this approach. An area of political research that has received a great deal of attention is to do with the study of media effects on public opinion. In many ways, this type of research lends itself well to the experimental setting of the laboratory, since the causal significance of media 'effects' on political attitudes is notoriously difficult to unravel. Most people are exposed to multiple, competing media messages each day, so it is very difficult to establish which messages matter and how. One particular aspect of media effects that has received a great deal of attention is the impact of negative campaigning by political parties, and, in particular, whether or not 'attack ads' have a demobilizing effect on voters during an electoral campaign.

One influential attempt to answer these questions was carried out by Stephen Ansolabehere and his colleagues. Through a set of innovative controlled experiments under laboratory

Table 8.2 Exposure to word choice and attitudes towards people from Syria

	Experimental group 1	Experimental group 2
Intervention	Migrants	Refugees
(1) Strongly disagree	11%	10%
(2) Disagree	27%	24%
(3) Agree	45%	44%
(4) Strongly agree	17%	23%
Mean	2.68	2.79

Question: Do you agree or disagree that Britain should allow more (migrants/refugees) from Syria to come and live in the UK?

conditions, Ansolabehere and colleagues (1994) examined the extent to which subjects who view negative campaign adverts are less likely to say they will vote than subjects who view more positive campaign adverts. In the study, subjects were randomly assigned to different groups. Subjects in each group then watched identical public information broadcasts. Each broadcast had an advertisement break which consisted of three adverts, and one of these adverts varied across the groups. In one experimental group, subjects were exposed to an advert which was political in content and negative in tone; in another experimental group, subjects were exposed to an advert which was political in content and positive in tone. Subjects in the control group were not exposed to a political advert, and just watched a product advert. The results of their study are reproduced in Table 8.3. Controlling for a variety of other factors that influence whether or not someone votes, Ansolabehere and colleagues found that people who are exposed to political adverts with positive political messages are about six percentage points more likely to vote than people who are exposed to negative political adverts. They thus conclude that 'attack' adverts have a demobilizing effect.

Given the random allocation of subjects to groups, any observed differences between the experimental and control groups can therefore be confidently attributed to experimental intervention. These types of experimental study therefore have a high level of internal validity. This means that we can be confident that the change in the outcome variable really was brought about by the key independent variable (the intervention), rather than some other factor. That is, it really is the variation in the tone of the political adverts that is responsible for the variation in willingness to vote. However, laboratory experiments are often criticized for lacking external validity. Since they take place in an 'artificial' environment, their findings cannot be easily generalized to the 'real world'. We might be confident that exposure to attack ads has a demobilizing effect in the laboratory setting, but does this effect also hold in the real world? Could attack ads be responsible (at least partly) for a decline in turnout?

Experimental laboratory research allows us to test hypotheses under controlled conditions designed to maximize internal validity. However, exercising a high degree of control over subjects can often lead to a reduction in the external validity of the findings—and, in particular, that variant of external validity that is called **ecological validity**. This means that, because they do not reflect a real-life situation, the findings of laboratory experiments may not be generalizable (or extended) to the 'real world'. When we take people out of their natural environment and study them in the laboratory, we are exerting some control over them. Consequently, we are possibly limiting how much we can generalize the findings to all people in natural settings. The laboratory study may therefore lack ecological validity because the controlled environment in which it is conducted is so unlike the real world that whatever results are obtained will be inapplicable to people in non-laboratory settings. The question, then, is: to what extent is a laboratory study true to life? Could it be replicated anywhere, using natural settings and conditions? Another reason why a laboratory study

Table 8.3 Exposure to negative campaigning and likelihood of voting

	Control group	Experimental group 1	Experimental group 2
Intervention	Product advertisement	Negative political advert	Positive political advert
Post-test (Y)	61% likely to vote	58% likely to vote	64% likely to vote

Source: Adapted from Ansolabehere and colleagues (1994).

may lack ecological validity is because of **reactivity**. Reactivity occurs when research study participants alter their behaviour as a result of being aware that they are participating in a study. Reactivity threatens ecological validity because, when it occurs, the results of the experiment might be *generalizable only to other people who are also being observed*.

To illustrate some of these points, we can return to Stephen Ansolabehere et al.'s study of 'attack ads'. One concern with the external validity of the findings might be that, since the behavioural and attitudinal consequences of exposure to the different types of campaign ads were measured by a survey conducted shortly after exposure occurred, it is unclear what the long-term impact of this was and how it might translate into actual electoral outcomes. If, for example, intention to vote declines immediately after exposure to negative advertising by three percentage points in the lab, does that imply that the mud-slinging senate campaign under study lowers actual turnout in the electorate by three percentage points? As Green and Gerber (2003: 101) argue, intention to vote is not the same thing as actual turnout; nor is one-time laboratory exposure the same thing as multiple-attempted exposures in the course of an actual campaign. Social media can also be an important tool in elections, as explored in Research in Action Box 8.1.

Field experiments

Although some scholars prefer laboratory experiments to field experiments because the lab offers the researcher tighter control over the treatment and how it is presented to subjects, others take the view that the generalizations from these types of study will be limited unless treatments are deployed, and outcomes assessed, unobtrusively in the field (Druckman et al. 2006). Because of concerns about the artificial environment of laboratory experiments,

Research in Action 8.1 A Sixty-One-Million-Person Experiment on Social Media and Political Mobilization

Can social media help to spread political engagement? And are people more likely to vote if they know their friends have voted? To answer these questions Robert Bond and colleagues (2012) carried out an experiment on sixty-one million Facebook users to find out. They carried out a randomized study on all Facebook users in the United States aged 18 years or above who accessed Facebook on the day of the 2010 US congressional elections. They randomly assigned users to one of three groups. The first 'social message group' was shown a statement at the top of their 'News Feed'. This message encouraged the user to vote, provided a link to find local polling places, showed a clickable button reading 'I Voted', showed a counter indicating how many other Facebook users had previously reported voting, and displayed up to six small randomly selected profile pictures of the user's Facebook friends who had already clicked the 'I Voted' button. The second 'informational message group' was shown the same information as the first group apart from the details about whether the user's Facebook friends had clicked the 'I Voted' button. The third 'control group' did not receive any information at the top of their feed. The researchers then matched 6.3 million users to publically available voter records to see if they did actually vote or not. From these official voter records, they found that people in the 'social message group' were more likely to vote than people in the 'informational message group' and people in the control group. This suggests that people are more likely to vote if they know that their friends have already voted, indicating that social influences can have an important impact on political mobilization.

political scientists have carried out experiments in real-world environments. These field experiments attempt to reproduce as closely as possible the conditions under which different political phenomena occur, thus increasing the external validity or generalizability of the findings. As Green and Gerber (2003: 94) observe, field experiments have two main strengths. First, random assignment ensures unbiased inference about cause and effect; and, second, the natural settings ensure that the results will tell us something useful about the real world. Field experimentation can therefore be an incredibly powerful tool for enabling researchers to draw unbiased and externally valid causal inferences about different social and political processes.

There is a long history in political research of field experimentation, or what Green and Gerber (2003) term controlled interventions, into the political world. They use the term 'controlled intervention' since many of the early studies did not assign subjects to treatment and control conditions on a purely random basis. However, in most other respects, they closely resemble field experiments. An early example of this kind of controlled intervention was carried out by Harold Gosnell (1927) on voter registration and turnout in Chicago prior to the 1924 and 1925 elections. Gosnell gathered the names, addresses, and background information of thousands of voting-age adults living in various Chicago neighbourhoods. He then divided these neighbourhoods into blocks, assigning (though not on a strictly random basis) certain blocks to the treatment condition of his experiment, which consisted of a letter urging adults to register to vote. Comparing the registration and voting rates in his treatment and control groups, Gosnell found his letter campaign to have produced a noticeable increase in political participation across a variety of ethnic and demographic groups.

A more recent and scientific example of this kind of experiment was carried out by Peter John and Tessa Brannan (2008) in Manchester prior to the 2005 British election. John and Brannan randomly selected the names of 6,900 people from the electoral register of a constituency in Manchester. The subjects were then randomly allocated to three groups (one control and two treatment groups). The researchers then selected one treatment group to receive a telephone call (the telephone group), and the other treatment group to receive a visit (the canvassing group). For both groups, the main purpose of the contact was to persuade the citizen to vote, both by providing reasons why it is important and by attempting to respond to any concerns about the voting process. The researchers had no contact with the control group. Comparing participation across the groups, John and Brannan found that turnout was significantly higher among the treatment groups than it was among the control group, with both telephone calls and canvassing having much the same effect on boosting turnout. The results of their study are reproduced in Table 8.4. In the control group turnout was 51.5 per cent, whereas turnout in the canvassing group who received face-to-face contact was 55.1 per cent and turnout in the telephone group who received a phone call was 55.0 per cent.

Table 8.4 Exposure to different types of canvassing and turnout

	Control group	Experimental group 1	Experimental group 2
Intervention	No contact	Face-to-face contact	Telephone contact
Post-test (Y)	51.5% voted in election	55.1% voted in election	55.0% voted in election

Source: Adapted from John and Brannan (2008).

These studies not only provide strong evidence about the causes of turnout in the real world, but they also provide strong evidence about what factors can increase turnout. Experimental field research can therefore be a very effective way of evaluating different policy initiatives and pilot programmes, since they provide a direct test of whether the initiative (or intervention) brings about a direct change in the outcome of interest. For this reason, field experiments can often be used to great effect in collaboration with political parties, policy-makers, or other agencies who are seriously interested in trying to understand the impact of some type of policy or intervention. For example, Wantchekon (2003) carried out a remarkable field experiment on clientelism in Benin during a national election, in which he was able to persuade political parties to randomize the types of appeal they made to voters in different villages (between programmatic appeals and patronage-based clientelist appeals). He was able to do this because there was considerable interest among the leading parties in Benin in learning about the effectiveness of alternative campaign strategies.

Wantchekon selected a number of villages and the inhabitants were then exposed to purely clientelist platforms and campaign strategies by one party and purely programmatic public policy platforms and appeals by the other party. The type of appeal that each party put forward varied between the different selected villages. The public policy message emphasized general policy goals, such as national unity and peace, eradicating corruption, alleviating poverty, developing agriculture and industry, and protecting the rights of women and children. The clientelist message emphasized specific promises made to the village for things like government jobs or local public goods, such as establishing a new local university or providing financial support for local workers.

Comparing these two experimental groups to the control group, where the platform was not manipulated and voters were exposed to the usual mixed platforms of the parties, Wantchekon found that parties which adopted the clientelist appeals tended to be much more successful than the parties which adopted the programmatic appeals, particularly for incumbent candidates. He also found that some groups of people responded to these appeals in different ways, and that women had a stronger preference for public goods messages than men.

Another great strength of experimental research is to try and uncover the way in which variables are related to each other when the direction of causality is uncertain. For example, a great deal of academic research has tried to unpick the relationship between newspaper readership and political attitudes and support. Does the newspaper someone reads influence their political preferences? Or do people's political preferences influence which newspaper they read? These so-called 'chicken-and-egg' problems can be very difficult to solve. Politicians certainly believe that newspapers can be very influential, as do the newspapers themselves. After the Conservative Party victory in the 1992 UK election, the *Sun* (a British tabloid) brazenly declared 'It's the Sun Wot Won it'. Yet hard empirical evidence to support this view is difficult to come by.

One way around the chicken-and-egg problem is to carry out experimental research. Gerber, Karlan, and Bergan (2009) report on an experiment they conducted during the 2005 Virginia gubernatorial election, designed to see if biased information sources affected voter behaviour. Washington has two national newspapers, the *Washington Post* and the *Washington Times*. Whereas the *Washington Post* is generally viewed as a liberal newspaper, the *Washington Times* is widely considered to be a more conservative paper. Gerber and

his colleagues selected a large sample of subjects about a month before the election, and then, after discounting any people who already read the *Post* or the *Times*, randomly assigned subjects to one of three groups. The first group received a free one-month subscription to the *Post*, the second group received a free one-month subscription to the *Times*, and the third group received neither. Subjects completed a survey at the beginning of the study (pre-test) and after the election (post-test). Gerber, Karlan, and Bergan (2009) found that subjects who had been assigned to the (liberal) *Washington Post* were eight percentage points more likely to vote for the (liberal) Democratic candidate than those not assigned a free newspaper. The results therefore provide evidence that political biases in newspapers can affect voting behaviour and political attitudes.

However, field experiments are not without their problems either, and in tackling the problem of external validity often associated with laboratory experiments, it is frequently argued that they introduce a new problem to do with internal validity. As we discussed in the previous section, when researchers exert a lot of control over a study (such as when they carry it out in a laboratory), it creates a degree of artificiality. This reduces the external validity of the experiment and makes it harder to generalize the findings. However, when researchers attempt to deal with this problem and carry out their experiments in the field, they have less control over the study and so can be less sure about the causal significance of the intervention. This can reduce the internal validity of the experiment. To illustrate this problem, we can return to the series of studies conducted by the psychologist Robert Rosenthal and his colleagues that we discussed in Chapter 3. Rosenthal and his colleague, Lenore Jacobson, conducted a study designed to determine the impact of teachers' expectations on student performance. The research question that Rosenthal and Jacobson (1968) addressed was: 'Does a teacher's expectations affect students' performance?'. Their hypothesis was that favourable expectations of teachers will lead to an increase in students' intellectual ability. The dependent variable was intellectual achievement (measured by IQ tests). The independent variable was teachers' expectations (those that resulted from the information communicated to them by the researchers). The researchers conducted a pre-test of their dependent variable: they had IQ tests administered to elementary school students (grades 1–6). They randomly selected 20 per cent of the students, who were assigned to serve as the experimental group; the remaining 80 per cent of the students represented the control group. The teachers were told that the students in the experimental group were academic 'bloomers', and would show unusual academic development over the coming year. At the end of the school year, Rosenthal and Jacobson retested all of the students. The IQs of students whom they had characterized as academic 'bloomers' (the experimental group) had statistically improved, and they had improved at a faster rate than those of the other children (the control group).

This study appears to meet the conditions of a valid experiment. In the ideal experimental design, we are able to hold constant all variables but one, and do so in a controlled, unchanging environment. But Rosenthal and Jacobson did not have complete control over the subjects and their environment, because the experiment occurred in a 'natural' setting. Consequently, student performance may have been affected by variables other than the teachers' expectations. Teachers may have introduced concrete changes that enhanced the learning environment of those students who had been characterized as 'intellectual bloomers', by, for instance, giving them extra or more difficult work, or choosing them to

undertake more enriching classroom or extracurricular activities. They may have communicated to parents that their children were more advanced, thus bringing about changes in home environments. There is thus sometimes thought to be a trade-off between ensuring high levels of internal validity and ensuring high levels of external validity.

In order to have confidence in the results of an experiment, it is important to have high levels of both internal validity and external validity. Given that internal validity can sometimes be a problem in field experiments and external validity can sometimes be a problem in laboratory experiments, some critics argue that, since both designs are flawed, there is not much point in using either method. But a more reasoned approach might be to try and incorporate both types of design together. For a study to really stand up, it should be possible to try and examine it in both settings. As Elinor Ostrom (2007: 26–7) says, 'To test theory adequately, we need to use methods that together combine external and internal validity. One gains external validity in doing field research, but internal validity in the laboratory.'

Natural experiments

The third main type of experimental design is often referred to as a natural experiment. In recent years the number of studies claiming to use this method has dramatically increased, and has more than tripled since the 1990s (Dunning 2012). The natural experiment relies on naturally occurring events as interventions rather than on interventions controlled by the researcher. Natural experiments are therefore 'discovered' by researchers rather than designed by researchers. According to Morton and Williams (2009), when natural interventions occur on a particular variable of interest, we can sometimes treat the intervention as if an experimentalist manipulated the variable. Even though in this case the researcher is not doing the intervening, the approach taken with the data is as if the researcher has done so. Such experiments have been utilized by scholars in a wide variety of fields, including political participation (Lassen 2005; Krasno and Green 2008; Dasgupta 2018), ballot design (Gordon and Huber 2007; Carman, Mitchell, and Johns 2008), political psychology (van der Brug 2001), ethnic politics (Abrajano, Nagler, and Alvarez 2005), comparative politics (Posner 2004), and bureaucracy (Whitford 2002).

The attraction of a natural experiment is that it removes many of the problems associated with carrying out laboratory and field experiments. First, since a natural experiment is naturally occurring, there is not the problem of artificiality associated with the laboratory experiment. And second, since the intervention occurs independently of the actions of the researcher, there are not the ethical issues which can create obstacles to doing field experiments. However, for a naturally occurring event to approximate a 'proper experiment', it is essential that the principles of experimental research are not violated. In particular, the condition of exogeneity is crucial. For a natural experiment to be valid, there must be variation in a causal variable that is independent of all other competing factors that may affect the outcome of interest. In practice, this is often difficult to achieve. Naturally occurring interventions are often not entirely random. For example, when a country changes its electoral system, we might think that this represents a good opportunity to examine the impact of institutional design on some aspect of political behaviour, such as turnout. We might even regard this as a natural experiment, since we have variation in our independent variable of interest (electoral system) while many other important factors appear to stay the same.

But a problem arises if not all factors do stay the same. For example, if the decision to change the electoral system is driven in part by concerns about citizen apathy and low levels of political support, then the variation in our variable of interest will be correlated with other important factors that drive turnout.

It is therefore important to try and ensure as far as possible that the principle of random allocation is not violated. Dunning (2012) distinguishes between natural experiments with true randomization and natural experiments with 'as-if' randomization. According to Dunning (2012: 48) true randomization occurs in situations in which an actual randomizing device with a known probability distribution assigns subjects to the treatment and control conditions. By contrast, in other situations, the assignment to treatment conditions might not be specifically generated by random allocation, but the researcher might be able to make a plausible claim that the allocation is as good as random. For example, Hyde (2007) examines the effect of international election observers on electoral fraud in Armenia, by comparing the electoral returns from polling stations which were 'observed' by election monitors with those from polling stations which were not. In order to make a compelling case that this study represents a natural experiment, Hyde argues that even though election monitors did not randomly select which polling stations to observe, their selections method was 'as-if' they had randomly selected polling stations since their method of selection was not informed by any criteria that could have predicted voting patterns. Hyde suggests that the presence of international election monitors reduced the share of the vote for the incumbent by just under six percentage points.

Sometimes the claim that treatment assignment is 'as-if' random is not particularly persuasive. In these situations the study in question is probably better characterized as an observational study than a natural experiment. Dunning (2012: chapter 8) provides a detailed discussion on how to evaluate whether treatment assignment is 'as-if' random or not. According to Dunning the whole point of randomization is that subjects should not be able to choose whether they go into the treatment or the control group. Any claim that treatment assignment is as good as random is therefore violated if subjects have the opportunity to self-select into treatment conditions. Similarly, the claim is violated if policy-makers or politicians assign subjects to different groups using a specific decision rule other than randomization.

True randomization often occurs when government officials make policy interventions in which the randomization of programme eligibility is an explicit part of the programme (Dunning 2012). Under these conditions, researchers may be able to exploit near experimental conditions to examine a wide variety of different issues. Dunning refers to these studies as 'lottery studies'. One such example is Angrist's (1990) study of the effects of military conscription on labour market earnings. This was a topic of some controversy after the Vietnam War. However, the observation that ex-soldiers earn less than non-soldiers is not really evidence to support the idea that military service damages future earnings, as the sort of people who become soldiers in the first place might be very different to those who don't. Angrist therefore used a natural experiment to try and sort out some of these confounding factors. From 1970 to 1972 the United States used a randomized lottery to draft soldiers for the Vietnam War. Young men were randomly assigned a number that ranged from one to 366 and all the men with a number below a given cut-off were eligible for the draft. Thus, young men at this time were in a sense randomly assigned to one of two groups: those who were eligible for the draft and those who weren't. Angrist found that those who were eligible for the draft tended to have somewhat lower future earnings than those who were not (see Table 8.5). Among white men eligible for the draft in 1971, average earnings in 1984 were

Table 8.5 Draft eligibility and future earnings

	Experimental group 1	Experimental group 2
Intervention	Eligible for the draft in 1971	Not eligible for the draft in 1971
Post-test (Y)	Average earnings in 1984: $15,813.	Average earnings in 1984: $16,172

Source: Adapted from Angrist (1990).

$15,813. By contrast, among white men not eligible for the draft in 1971, average earnings in 1984 were slightly higher at $16,172. Thus, assignment to draft eligibility in 1971 caused an estimated decrease in future yearly earnings of $359, or just over two percentage points (see Dunning 2012: 8, for further discussion of this example).

Another example of a study based on a lottery (in more than one way) was carried out by Doherty, Gerber, and Green (2006), who were interested in the effect of income on political attitudes. They carried out a survey on 342 people who had won money on the lottery between 1983 and 2000 in an Eastern US state, and asked a series of questions about taxes, redistribution, and other social and economic policies. The political attitudes of lottery winners obviously cannot be meaningfully compared to the population at large, since those people who play the lottery are different in a number of respects to those who don't. However, among those people who play and win the lottery, the amount of winnings are randomly assigned. By separating income from education, class background, and other factors, lotteries can in principle enable researchers to isolate the effects of affluence on political attitudes (Doherty, Gerber, and Green 2006: 455). Doherty and colleagues were able to use this information to compare the effect of larger and smaller lottery winnings on political attitudes, as in a sense lottery winners were randomly assigned to one of these two groups. They find that lottery winnings have a substantial effect on views towards estate taxes, which touch directly on respondents' material interests, and a marginally significant effect on views towards income redistribution.

Researchers have also used natural experiments to examine the impact of different ballot designs on vote choice. Miller, Krosnick, and Lowe (1998) examined the effects of ballot order (the order in which candidates' names appear on the ballot) on votes for political candidates in the United States, by exploiting the fact that, in certain Ohio counties, candidates' names are rotated from one precinct to the next. Miller and colleagues found that candidates at the top of the ballot win an average vote share of 2.5 percentage points more, with the largest effects turning up in contests without an incumbent contestant and where candidates' names appeared without party affiliations. In an 'unfortunate' natural experiment, Carman, Mitchell, and Johns (2008) examined the effects of ballot design and voting instructions on the number of spoiled votes (ballots which had not been correctly filled in) in Scotland, by exploiting the fact that two different ballot designs were used in the Scottish Parliamentary elections of 2007 (one which provided full instructions and one which provided abbreviated instructions). Carman and colleagues found that, even after taking into account other potentially important factors, such as social deprivation, the predicted number of spoiled votes was 65 per cent greater for ballots with abbreviated instructions. Carman and colleagues call the experiment 'unfortunate' because the decision to abbreviate the ballot instructions was not an intentional one, and only came to light after the election, when academics and the media began to investigate why there were so many spoiled ballots.

Natural experiments can also arise out of specific policy interventions. In an unusual natural experiment from India, Rikhil Bhavnani (2009) examined the impact of electoral

quotas on women's representation in the *Panchayat* (village council). In particular, Bhavnani examined whether electoral quotas for women alter women's chances of winning elections even after the quotas are withdrawn. He was able to answer this question by exploiting the fact that randomly chosen seats in local legislatures are set aside for women for one election at a time. Bhavnani found that the probability of a woman winning office in a seat which had previously been reserved for women (but was not any more) was five times higher than if the constituency had not been previously reserved for women.

Issues in experimental research

Despite the obvious strengths of experimental research for testing causal hypotheses, and despite the many examples of interesting and valuable research that have been carried out using experimental methods, experimental designs are still much less common in the social sciences than in the natural sciences. Indeed, experimental research is often viewed with a great deal of suspicion, and is sometimes thought to work better in theory than in practice. This view is neatly summed up by Arendt Lijphart: 'The experimental method is the most nearly ideal method for scientific explanation, but unfortunately it can only rarely be used in political science because of practical and ethical impediments' (1971: 684–5).

The practical and ethical obstacles to experimental research have received a great deal of attention. One of the main practical objections is that, with respect to many key areas in politics, researchers simply cannot randomly assign subjects to groups and manipulate the independent variable in order to measure their impact on the dependent variable. It is one thing to randomize whether or not voters are exposed to campaign material, but it is another thing to randomize the foreign policies of governments. The really big political science variables, like culture, electoral system, level of economic development, ethnic heterogeneity, cannot be manipulated. As Green and Gerber (2003) note, it is difficult to imagine how one could randomly assign presidential and parliamentary regimes for the purpose of evaluating their relative strengths and weaknesses. Similarly, it is simply not possible to go around randomizing the determinants of war to see which given set of factors is most likely to cause two countries to start a conflict. For these reasons, it is commonly thought that political research can never fully embrace experimental methods, and that it will remain something of a niche method. This view is put forward by Smith (2002), who suggests that experimental research can only be conducted on a relatively small (and minor) fraction of the political questions that most interest people.

Although there is some truth to this criticism, it is perhaps wise not to overstate it. With a bit of imagination, the prospects for innovative field experiments are much greater than is commonly assumed. Experimental research has been carried out on big issues, such as crime and security (Keizer, Lindenberg, and Steg 2008), clientelism (Wantchekon 2003), police discrimination (Heussenstamm 1971), and the influence of the media (Gerber, Karlan, and Bergan 2009). Moreover, as Green and Gerber (2003) argue, there is a balance that political researchers need to strike between answering 'big questions' badly and 'small questions' well. As they put it:

> If we think of the expected value of research as being the product of the intrinsic value of a research question times the probability that knowledge will be advanced by the evidence flowing from that research, this trade-off comes into sharper focus.

Experimental research can therefore be an extremely effective method for helping to answer narrow, tractable questions.

There are also ethical issues that must be considered when conducting experiments. Ethical issues are relevant to all types of social inquiry, and we address them in detail in Chapter 6, but they also have special significance for experimental research. As Gerry Stoker (2010) points out, ethical issues tend to be more pronounced in experimental research than in purely observational research. The act of intervening to change something raises more ethical issues than simply observing what is happening, particularly if that intervention causes harm or distress to participants or relies on some sort of deception. A seminal psychology laboratory experiment carried out by Stanley Milgram (1963) is probably the most famous example of this kind of deception. It tested the willingness of participants to harm another person 'while only following orders' by getting them to administer electric shocks to others who appeared to be suffering pain. The study raises a number of ethical issues. The first is to do with deception: participants were lied to. They did not really administer electric shocks. The second is that it may have caused participants distress: they may have been upset to learn that they were capable of basically killing someone just because a person in a position of authority told them to do so. However, these ethical challenges are not insurmountable; and there are a number of strategies that can be employed to try and deal with them (see Box 8.1).

Related to ethical issues, there are also issues of professional conduct and practice that need to be considered when carrying out experiments and reporting the results. As discussed in Chapter 7, there have been a number of high-profile cases where it emerged that researchers fabricated field experiments and basically made up their data. This kind of professional misconduct is clearly wrong. Although it does not happen very often, when it does happen it can do a great deal of reputational damage to the discipline and make people

BOX 8.1 Ethical Issues in Experimental Research

Ethical issues are extremely important in experimental research, as almost by definition experiments involve some sort of deception. In laboratory experiments subjects may know that they are participating in a research activity, but may not know that they are participating in an experiment (or may not know what the experiment is about). In field experiments subjects may not even know that they are participating in a research project. This makes the issue of informed consent very challenging. Researchers in a number of experiments which were carried out on subjects who did not directly consent to be included in the study have justified their approach by saying that (1) no harm was done to participants, and (2) the interventions that were administered were very similar to information (or stimuli) that participants were already receiving. Most universities and research institutions now have a system where all prospective experiments must be reviewed, discussed, and approved by an Ethical Review Board (or Institutional Review Board). The job of these boards is to assess the risks and benefits of each study, and to weigh the balance between risk to the subject and benefit to society. When deception is used, review boards are particularly careful to make sure that the deception is absolutely necessary. One reason why it may be deemed necessary from a research point of view is when awareness of the hypotheses being tested might alter participants' behaviour, thus biasing the results. At the end of the experiment researchers should explain to subjects what the deception was and why deception was deemed necessary for the research. However, some people disagree with this approach and argue that any deception is unethical and can never be justified.

less likely to trust the results of experiments. The principles of replication (repeating experiments on a new sample) and verification (making the data from experiments publicly available so they can be re-analysed) can therefore help to make the research process more transparent, which in turn gives us confidence in the results as it adds an extra layer of scrutiny to how the data has been collected and analysed.

A slightly different aspect of professional practice is what is known as publication bias. This arises when experiments that yield 'significant results' are more likely to be published than experiments that yield 'null results'. This can create a problem when trying to assess the state of knowledge in a field or on a particular topic, since null results are largely unobservable to the scholarly community (Franco, Malhotra, and Simonovits 2014). In a systematic study of publication bias, Franco et al. (2014) examined 221 social science experiments that had received funding in the USA. Because the studies had all undergone rigorous peer review to decide which got funded, all of the studies in the sample had passed a substantial quality threshold. However, those experiments which produced strong results were 40 per cent more likely to be published. This was largely because scholars did not bother to write up the results from experiments that did not produce the desired results, partly because they didn't think the results would get published and partly because they lost heart in 'unsuccessful' projects. However, null results are just as important as significant results and can be just as interesting, so it is important that they are reported. As we often tell students, as long as the research question is interesting and the hypothesis that is being tested is interesting, then the results—whether they confirm the hypothesis or not—should also be interesting. In this sense the important thing is to provide a thorough test of the hypothesis under investigation, rather than merely to find support for the hypothesis.

Conclusions

Experimental research has a number of strengths. It allows researchers to investigate causal hypotheses with a great deal of confidence. There is no other research design that is quite so effective at establishing causal connections, and the principles that underpin experimental research inform much of what we do in political research. Even if experimentation is not used widely in political research (though this is beginning to change), most of our research designs are, in effect, an effort to approximate the logic of experimental design as closely as possible. In particular, the ideas of experimental control permeate much of what we do, even if it is dealt with in different ways. To compensate for our inability to completely control any one case, researchers in our field can study many cases through the use of statistical analysis (see Chapters 15 to 17). We can also try and control for important factors using comparative research (see Chapter 9).

But experimental research also has a great deal of practical and, specifically, policy relevance. Experimental research can be a very powerful tool for monitoring and assessing the implementation of different policy actions. There is certainly a great deal of potential for academics and policy-makers to work more closely together to develop experimental studies. And, as Stoker (2010) suggests, pursuing this strategy would really help to make political research more relevant to politics, political actors, and citizens. The benefit for academics is that they are able to collect data to test hypotheses that they would not otherwise

be able to investigate. For example, Wantchekon would not have been able to carry out his study on clientelism in Benin without the support of the main political parties who participated in his experiment. The benefit for policy-makers is that they will get valid and reliable evidence on what difference, if any, their policy intervention makes. Policies can be very expensive to implement and can also be very ineffective at bringing about the desired change. New policies thus provide a ripe opportunity for experimental research. Indeed, just as new medicines and drugs are required to undergo rigorous trial before they are approved, there is a lot to be said for adopting the same standards for policy initiatives.

The prospects for experimental research therefore look bright. Experiments provide a rigorous and scientific way to test causal hypotheses. They also offer the potential to engage with policy-makers in order to assess the effectiveness of new policies. With a little imagination, experiments can be used to answer a wide variety of questions which tackle a range of theoretical and empirical puzzles.

Questions

1. What are the strengths of experimental research?
2. Why is the experimental approach not used very often in political research?
3. What are the practical and ethical obstacles to experimental research? How can these obstacles be overcome?
4. Is deception ever justified? If so under what circumstances?
5. Experimental research can only be used to answer small questions which are of minor interest to most people. Do you agree?
6. What kind of research questions may be particularly well-suited to experimental designs?
7. What are the relative strengths and weaknesses of laboratory and field experiments?

Guide to Further Reading

Druckman, J. N., D. P. Green, J. H. Kuklinski, and A. Lupia (eds) (2011), *Cambridge Handbook of Experimental Political Science* (Cambridge: Cambridge University Press).

This book contains a collection of thirty-six articles on experimental research in political science, covering topics such as internal and external validity, and using students as subjects, as well as detailed chapters on laboratory experiments, field experiments, survey experiments, and more. The collection also includes a number of chapters that use experiments in practice to answer questions of substantive and theoretical interest.

Dunning, T. (2012), *Natural Experiments in the Social Sciences: A Design-Based Approach* (Cambridge: Cambridge University Press).

This book provides an accessible and sophisticated guide to natural experiments in the social sciences. It provides rich examples of successful and not so successful experiments, and includes detailed guidance on how to find or discover natural experiments, as well as how to evaluate natural experiments, particularly with reference to claims that are made about random assignment. It is essential reading for anyone interested in knowing more about experimental research.

Morton, R. B. and K. C. Williams (2009), *From Nature to the Lab: The Methodology of Experimental Political Science and the Study of Causality* (Cambridge: Cambridge University Press).

This book provides a detailed discussion of the different ways in which experimental designs can be used in political research. It provides a comprehensive account of the different issues researchers must engage with, and is full of interesting examples to illustrate their points.

Stoker, G. (2010), 'Exploring the Promise of Experimentation in Political Science: Micro-Foundational Insights and Policy Relevance', *Political Studies* 58: 300–19.
This article examines the ways in which experimental research can be used in mainstream political science and the potential the approach has to engage with policy-relevant research.

References

Abrajano, M., J. Nagler, and R. Alvarez (2005), 'Race Based vs. Issue Voting: A Natural Experiment', *Political Research Quarterly* 58(2): 203–18.

Angrist, J. (1990), 'Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records', *American Economic Review* 80(3): 313–36.

Ansolabehere, S. and S. Iyengar (1997), *Going Negative: How Political Advertisements Shrink and Polarize the Electorate* (New York: Free Press).

Ansolabehere, S., Iyengar, S., Simon, A., and Valentino, N. (1994), 'Does Attack Advertising Demobilize the Electorate?', *American Political Science Review* 88(4): 829–38.

Bhavnani, R. (2009), 'Do Electoral Quotas Work After they are Withdrawn? Evidence from a Natural Experiment in India', *American Political Science Review* 103(1): 23–35.

Bond, R., C. Fariss, J. Jones, A. Kramer, C. Marlow, J. Settle and J. Fowler (2012), 'A 61-million-person experiment in social influence and political mobilization' *Nature* 489: 295–98

Carman, C., J. Mitchell, and R. Johns (2008), 'The Unfortunate Natural Experiment in Ballot Design: The Scottish Parliamentary Elections of 2007', *Electoral Studies* 27(3): 442–59.

Dasgupta, A. (2018), 'Technological Change and Political Turnover: The Democratizing Effects of the Green Revolution in India', *American Political Science Review* 112(4): 918–38.

de Vaus, D. (2001), *Research Design in Social Science* (London: Sage).

Doherty, D., A. Gerber, and D. Green (2006), 'Personal Income and Attitudes toward Redistribution: A Study of Lottery Winners', *Political Psychology* 27(3): 441–58.

Druckman, D. (1994), 'Determinants of Compromising Behavior in Negotiation: A Meta Analysis', *Journal of Conflict Resolution* 38: 507–56.

Druckman, J. and C. D. Kam (2011), 'Students as Experimental Participants: A Defense of the "Narrow Data Base"', in J. N. Druckman, D. P. Green, J. H. Kuklinski, and A. Lupia (eds), *Cambridge Handbook of Experimental Political Science* (Cambridge: Cambridge University Press).

Druckman, J., D. P. Green, J. H. Kuklinski, and A. Lupia (2006), 'The Growth and Development of Experimental Research in Political Science', *American Political Science Review* 100(4) (November): 627–35.

Dunning, T. (2012), *Natural Experiments in the Social Sciences: A Design-Based Approach* (Cambridge: Cambridge University Press).

Franco, A., N. Malhotra, and G. Simonovits (2014), 'Publication Bias in the Social Sciences: Unlocking the File Drawer', *Science* 345(6203): 1502–5.

Fréchette, G., J. Kagel, and M. Morelli (2005), 'Behavioral Identification in Coalitional Bargaining: An Experimental Analysis of Demand Bargaining and Alternating', *Econometrica* 73(6): 1893–937.

Frohlich, N. and J. A. Oppenheimer (1992), *Choosing Justice: An Experimental Approach to Ethical Theory* (Berkeley, CA: University of California Press).

Gerber, A. and D. Green (2000), 'The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment', *American Political Science Review* 94(3) (September): 653–63.

Gerber, A., D. Karlan, and D. Bergan (2009), 'Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions', *American Economic Journal: Applied Economics* 1: 35–52.

Geva, N. and A. Mintz (1997), *Decision-Making on War and Peace: The Cognitive-Rational Debate* (Boulder, CO: Lynne Rienner Publishers).

Gordon, S. C. and G. A. Huber (2007), 'The Effect of Electoral Competitiveness on Incumbent Behavior', *Quarterly Journal of Political Science* 2(2): 107–38.

Gosnell, H. F. (1927), *Getting-out-the-vote: An Experiment in the Stimulation of Voting* (Chicago: University of Chicago Press).

Green, D. and A. Gerber (2003), 'The Underprovision of Experiments in Political Science', *Annals of the American Academy of Political and Social Science* 589: 94–112.

Habyarimana, J., M. Humphreys, D. Posner, and J. M. Weinstein (2007), 'Why Does Ethnic Diversity Undermine Public Goods Provision?', *American Political Science Review* 101: 709–25.

Heath, O. and Tillin, L. (2018) 'Institutional Performance and Vote Buying in India', *Studies in Comparative International Development* 53 (1): 90–110.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis (eds) (2004), *Foundations of Human Society* (Oxford: Oxford University Press).
Heussenstamm, F. K. (1971), 'Bumper Stickers and the Cops', *Society* 8: 32–3.
Hyde, S. (2007), 'The Observer Effect in International Politics: Evidence from a Natural Experiment', *World Politics* 60(1): 37–63.
John, P. and T. Brannan (2008), 'How Different are Telephoning and Canvassing? Results from a "Get Out the Vote" field experiment in the British 2005 General Election', *British Journal of Political Science* 38: 565–74.
Keizer, K., S. Lindenberg, and L. Steg (2008), 'The Spreading of Disorder', *Science* 322: 1681–5.
Krasno, J. and D. Green (2008), 'Do Televised Presidential Ads Increase Voter Turnout? Evidence from a Natural Experiment', *Journal of Politics* 70(1): 245–61.
Lassen, D. (2005), 'The Effect of Information on Voter Turnout: Evidence from a Natural Experiment', *American Journal of Political Science* 49: 103–18.
Lijphart, A. (1971), 'Comparative Politics and the Comparative Method', *American Political Science Review* 65(3): 682–93.
Lodge, M., K. M. McGraw, and P. Stroh (1989), 'An Impression-Driven Model of Candidate Evaluation', *American Political Science Review* 83: 399–419.
Mack, V. and L. Stoetzer (2019), 'Election Fraud, Digit Tests and How Humans Fabricate Vote Counts—An Experimental Approach', *Electoral Studies* 58: 31–47.
Milgram, S. (1963), 'Behavioral Study of Obedience', *Journal of Abnormal and Social Psychology* 67(4): 371–8.
Miller, J. M., J. A. Krosnick, and L. Lowe (1998), 'The Impact of Candidate Name Order on Election Outcomes', *Public Opinion Quarterly* 62: 291–330.
Morton, R. B. and K. Williams (1999), 'Information Asymmetries and Simultaneous versus Sequential Voting', *American Political Science Review* 93: 51–67.
Morton, R. B. and K. Williams (2008), 'Experimentation in Political Science', in J. Box-Steffensmeier, D. Collier, and H. Brady (eds), *The Oxford Handbook of Political Methodology* (Oxford: Oxford University Press).
Morton, R. B. and K. Williams (2009), *From Nature to the Lab: The Methodology of Experimental Political Science and the Study of Causality* (Cambridge: Cambridge University Press).
Ostrom, E. (2007), 'Why Do We Need Laboratory Experiments in Political Science?' Paper presented at the 2007 American Political Science Association annual meeting, Chicago, IL.
Posner, D. (2004), 'The Political Salience of Cultural Difference: Why Chewas and Tumbukas are Allies in Zambia and Adversaries in Malawi', *American Political Science Review* 98(4) (November): 529–45.
Rosenthal, R. and L. Jacobson (1968), *Pygmalion in the Classroom* (New York: Holt, Rinehart & Winston).
Simon, A. and T. Sulkin (2001), 'Habermas in the Lab: An Experimental Investigation of the Effects of Deliberation', *Political Psychology* 22: 809–26.
Smith, R. M. (2002), 'Should We Make Political Science More of a Science or More about Politics?', *PS: Political Science and Politics* 35(2) (June): 199–201.
Song, B. K. (2019), 'Misleading Ballot Position Cue: Party Voting in Korea's Nonpartisan Local Elections', *Electoral Studies* 58: 1–11.
Stoker, G. (2010), 'Exploring the Promise of Experimentation in Political Science: Micro-Foundational Insights and Policy Relevance', *Political Studies* 58: 300–19.
Townsley, J. and S. Turnbull-Dugarte (2019), 'Can Parties Recruit Postal Voters? Experimental Evidence from Britain', Electoral Studies. https://doi.org/10.1016/j.electstud.2019.02.013
van der Brug, W. (2001), 'Perceptions, Opinions and Party Preferences in the Face of a Real-World Event: Chernobyl as a Natural Experiment in Political Psychology', *Journal of Theoretical Politics* 13(1): 53–80.
Wantchekon, L. (2003), 'Clientelism and Voting Behavior', *World Politics* 55: 399–422.
Wattenberg, M. P. and C. L. Brians (1999), 'Negative Campaign Advertising: Demobilizer or Mobilizer?' *American Political Science Review* 93: 891–900.
Whitford, A. (2002), 'Decentralization and Political Control of the Bureaucracy', *Journal of Theoretical Politics* 14: 167–93.

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

9 Comparative Research

Chapter Summary

This chapter examines the principles of comparative research design and discusses the issues and problems associated with different aspects of the approach. In doing so, we pay special attention to the issue of case selection, the common sources of error that are associated with comparative research, and what can be done to try and avoid or minimize them. We examine:

- case selection;
- case study;
- small-N comparison;
- large-N studies;
- selection bias;
- measurement.

Introduction

This chapter examines the principles and uses of comparative research. The comparative method is one of the most widely used methods in political research and is frequently used to study a wide range of political phenomena, from democratization to civil war to institutional design to public policy. The scope for comparative research is almost limitless. However, often the term 'comparative politics' is used rather vaguely to simply mean the study of 'foreign' countries outside the UK and USA (van Biezen and Caramani 2006). But the term 'comparative methods' really implies something more than this. As Arend Lijphart (1971) wrote, the term 'comparative politics' indicates the *how* of the analysis, but not the *what*. Comparative politics is first and foremost a methodological approach, rather than an area of substantive interest. What unites and defines comparative studies therefore has little to do with the sort of questions asked, but rather with the method used. Following Peter Mair (1996) and van Biezen and Caramani (2006), we understand the comparative method primarily in terms of the rules and standards and procedures for identifying and explaining differences and similarities between cases (often, but not always, defined in terms of countries), using concepts that are applicable in more than one case or country.

There are many different ways in which this can be done. Indeed, the comparative method actually involves a number of different methods. These methods can be distinguished primarily in terms of how many countries (or cases) are compared (denoted by

the letter **N**), and how the cases for analysis are selected. Both aspects of case selection are very important. Broadly speaking, there are three main approaches. There are large-N studies (involving the analysis of many cases), small-N studies (involving the analysis of a small number of cases, typically 2, 3, 4, but with no real upper limit) and single-N studies (otherwise known as case studies). In this chapter we discuss these different comparative approaches, provide an overview of the methodological issues that they raise, and evaluate the strengths and weaknesses of each approach.

The principles of comparison

Comparison, either directly or indirectly, forms an essential part of almost all empirical analysis. In a sense it is impossible not to compare. We do it all the time, whether we are aware of it or not. Any discussion of cause and effect relies on comparing the presence of one factor against the absence of another. This has led some scholars to define the comparative method rather broadly. For example, Gabriel Almond (1966) equates the comparative method with the scientific method generally, and says 'it makes no sense to speak of a comparative politics in political science since if it is a science, it goes without saying that it is comparative in its approach'. According to this understanding, all empirical research is comparative. We might even regard the statistical analysis of survey data as a form of comparative analysis, where the **unit of analysis** is an individual. For example, we can compare young people with old people to see if there is an association between age and turnout (see Chapter 16). But the comparative method more usually refers to the study of social and political *aggregates* rather than individuals. In particular, comparative politics is frequently based on comparing differences (or similarities) between countries. But it can also be used to compare differences between units within countries, such as regions, organizations, political parties, pressure groups, or whatever.

Like all methods, comparative methods are only as useful as the purpose to which they are put. If the research project has clear theoretical goals, then comparative research can be a valuable tool for helping to answer questions about different political phenomena. But if there is little theoretical justification for the comparison, or the justification is not made clear, then comparative research is unlikely to be of much use in developing or testing theory. It is very important then to think carefully about the purpose of the comparison, and to avoid falling into the trap of just comparing different countries for the sake of it, without any clear theoretical goals guiding the selection of countries to analyse.

Broadly speaking, comparative methods can be used in three main ways: (i) to apply existing theory to new cases; (ii) to develop new theory or hypotheses; and (iii) to test theory.

One of the key strengths of comparative research is that it helps to broaden our intellectual horizons. We can use comparison to see if what we think is a self-evident truth in one context also works in the same way in a different context. Comparison therefore has a number of advantages over the single country case study, and helps us to guard against the twin dangers of what Rose (1991) has labelled *false uniqueness* and *false universalism*. False uniqueness emphasizes the specificity of the case, entirely ignoring the general social forces at work, and does not move beyond 'thick description'. Problems to do with false uniqueness can sometimes be found in area studies, where researchers emphasize how unique—or

exceptional—their chosen country of analysis is, and seal themselves off from wider engagement with what is being written about in other countries. By contrast, false universalism assumes that the theory tested in one country/context will be equally applicable to other countries. For example, what is found to matter in the United States, whether it is the importance of 'resources for political participation', 'the association between social diversity and trust', or 'the impact of lobby groups on government policy' is often assumed to refer to universal truths about the way politics works, rather than context-specific findings particular to one country. The only way we can ever establish uniqueness or universalism is through comparison.

An important function of comparison is therefore descriptive. Do the same things work in the same ways across different contexts? We can use comparison to apply theory developed in one context (or **case**) to another context. Rokkan (1966: 19) refers to this as a 'micro replication', which is designed 'to test out in other national and cultural settings a proposition already validated in one setting'. This type of approach is often described as theory-infirming or theory-confirming (Lijphart 1971). For example, one influential theory in studies on political behaviour which emerged in the United States is that rich people and well-educated people tend to be more politically active because they possess the 'civic skills' that reduce some of the obstacles to participation (Verba, Schlozman, and Brady 1995). This theory is not framed in a context-specific way that is peculiar only to the United States (where the research was carried out), and so might be thought to apply across different contexts and be applicable to other countries. One useful purpose of comparison is therefore to try and establish whether this is the case or not, and whether the general theory proposed really is a general theory or not (false universalism). Subsequent research carried out in different countries has found that there is in fact substantial variation in terms of how important these so-called 'civic skills' are for political participation. In Britain, for example, although they appear to be important, they are less important than in the US (Parry, Moyser, and Day 1992). And further afield, in India, they are not important at all, and, if anything, it is those who are the most disadvantaged in society who tend to participate the most (Yadav 2000; Michelutti 2008).

Comparison can therefore serve a useful descriptive purpose, and be used to map the extent to which different theories apply (or can be replicated) across different contexts, and the extent to which different social and political phenomena will occur across different countries. But at the same time descriptive comparison only takes us so far. It is one thing to show that there is variation across contexts, or that some accounts of political participation work better in one country than in another; it is another thing to try and develop and test theories in order to explain these differences.

It is to this end that comparative methods are most frequently used. Comparison can be used to try and generate and test hypotheses about the factors that explain these variations. The advantages of comparative cross-national analysis are manifold. First, comparative methods can be used to merge different levels of analysis, and link international, national, and domestic factors in order to explain a particular political phenomenon. For example, by comparing individuals nested within countries, we are able to examine the impact of country-level factors (such as institutional arrangement) on individual behaviour. For example, comparative research on voter support for extreme right-wing parties shows that whether or not someone votes for the far right is influenced both by individual factors,

such as a person's attitudes towards immigration and national factors, such as whether the person lives in a country that uses proportional representation or not. Without comparison, we would not be able to examine the importance of these 'aggregate level' factors, and so our explanation of a particular phenomenon would be partial and underdeveloped. Comparison thus allows the researcher to move beyond a purely individualistic explanation of political phenomena.

How we compare

The comparative method is primarily a method for case selection. Do we compare many or few cases? And if we compare few cases, then on what basis do we select them? In the following sections we discuss the three main forms of comparison (case, small-N, and large-N studies). In particular, we focus on how we select cases for analysis, the different ways in which this can be done, and some of the problems and trade-offs that are involved with each approach. Whereas small-N studies are better suited to generating hypotheses than testing hypotheses, large-N studies are able to do both.

Case study

The great advantage of the case study is that by focusing on a single case, that case can be intensively examined. For this reason, case studies remain one of the main forms of research in Comparative Politics (Geddes 2003; George and Bennett 2005). On the face of it, it might appear to be a contradiction in terms to talk about a comparative case study, since it is not immediately obvious where the comparison comes from if there is only one case involved. But good case studies are nearly always situated in a comparative context. They address theory or issues that have wider intellectual relevance, use concepts that are applicable to other contexts, and may even seek to make inferences that apply to countries beyond that of the original case. For example, Arend Lijphart's (1977, 1999) studies of consociationalism in the Netherlands, Robert Putnam's (1993) study of social capital in Italy, and Paul Brass's (2003) study of ethnic violence in India are all regarded as classic 'comparative studies', even though each is only based on the analysis of a single country. What sets them apart is that they have all managed to develop arguments and theories that are relevant to many other contexts, and not only to say something meaningful and interesting about the case in question, but also to say something meaningful about general political phenomena.

Good case studies therefore possess two important characteristics. The first is that they say something interesting and meaningful about the case that is being studied. For example, a case study of ethnic violence in India should help to shed light on the sources of conflict in India, and contribute to the academic literature that has been written on the subject with reference to the particular case. That is, the findings of the study should be *internally valid*. However, a good comparative case study should also aim to say something more general, and engage with wider academic debates that might be applicable to other contexts and other cases. For example, does the study only shed light on the sources of ethnic violence in India, or could the implications of the argument and analysis also be relevant for helping

to explain the sources of ethnic conflict in other parts of the world, such as in Africa? This involves setting the case in comparative context, and proposing theories or explanations that are also *externally valid* (at least hypothetically).

Although there is a widespread assumption that case studies represent a type of qualitative research, this is not strictly true. Case studies can be based on a wide variety of different data-gathering strategies, such as interviews, surveys, ethnography, focus groups, historical documents, policy documents, and speeches, to name just a few. Indeed, many single-country case studies are based on the quantitative analysis of survey data, for example on voting behaviour in Britain, or public attitudes towards the environment in Germany, or public attitudes towards immigrants and ethnic minorities in the Netherlands. In the following chapters, we will explore some of these methods of data collection in more detail, so the focus here is more concerned with issues of case selection.

It is the extent to which the case study engages with wider comparative themes that makes it comparative. It can do this in one of two main ways. First, it can apply theory developed in one context to another context in order to assess whether the original theory 'works', thus addressing issues to do with false uniqueness and universalism. Or, second, it can seek to develop a new theory, and generate hypotheses that can be applied to other cases. However, because the case study is only based on the analysis of a single case, a case study cannot really be used to directly test theory. For this reason, we should always treat the inferences made from a case study with a certain degree of caution. They may sound plausible, but until they have been tested on a large number of countries, we have no way of knowing if they are.

These two main types of case study can be used in a variety of ways (Lijphart 1971; Gerring 2004, 2007a; Yin 2009). Broadly speaking, we can distinguish between case studies that (i) provide descriptive contextualization; (ii) apply existing theory to new contexts; (iii) examine exceptions to the rule; and (iv) generate new theory. Issues to do with case selection and what makes a good country to study depend upon the purpose of the case study. You must provide a rationale for why the specific case or set of cases you selected, from among all those in the larger population, were chosen. Researchers may select cases because they are *critical* (to testing a theory), or *revelatory* (reveal relationships which cannot be studied by other means), or *unusual* (throw light on extreme cases) (Yin 2009).

One function of the case study is contextual description. Purely descriptive case studies do not seek to advance or apply theory, but rather to provide a thick description of a particular event or phenomenon in a particular country. These types of study are most closely associated with ethnography (see Chapter 13) and can be a valuable source of data for secondary analysis. Given that these types of study do not engage with theoretical literature, it does not make much difference how the cases are selected for analysis, since there is no attempt made to use the case to make any wider inferences. However, this approach is not used very often.

More commonly, case studies are used to *apply theory* from one context to see if it still holds in another context. Research in Politics and International Relations is full of examples of this type of research. In the field of electoral behaviour, many concepts have been exported from research on the United States to other countries. Perhaps one of the most important of these 'exports' is the concept of party identification or party attachment, which states that some voters develop an attachment to a political party which is so strong that it forms a part of their identity (like a religious, national, or ethnic identity) and in turn shapes the way in which they interpret political information and events. This concept was

developed in the United States in the 1950s by Angus Campbell and his colleagues at the University of Michigan, and was then applied to the British context by Butler and Stokes in the 1960s (see Campbell et al. 1960; Butler and Stokes 1969). It has since been applied to many other countries in order to explain a wide variety of political behaviour.

Case studies are an incredibly powerful tool for examining whether concepts and theories travel, and whether (or not) they work in the same way in cases other than where they were originally developed. When choosing a case study to analyse for these purposes, Geddes (2003) suggests two main criteria for case selection. The first is that the case should be representative of the domains of the theories they are intended to test. That is, the case should provide a fair test of the theory. One must therefore identify the universe of cases to which the theory (or hypothesis) applies. The second criterion is that cases used for testing arguments should be different from the cases from which the arguments were induced. In this way, each case study provides a new test of the theory, and contributes to the gradual accumulation of knowledge in favour (or against) the theory.

Case studies can be used to examine specific **outliers** or **deviant cases**, and to examine countries (or cases) that do not fit existing theory and are known to deviate from established generalizations. This type of case study is still based on developing existing theory, and seeks to uncover the reasons why the theory does not apply to the case in question. Case studies of this type have often focused on countries that are thought to be 'exceptional' for one reason or another, such as why there is no socialist political party in the United States, even though experience of industrialization suggests there should be one, or why there is no democracy in rich countries like Saudi Arabia, when modernization theory suggests there should be. Rogowski (2004: 82) notes the central importance of a deviant case in which the outcome is unexpectedly wrong: 'A powerful, deductive, internally consistent theory can be seriously undermined … by even one wildly discordant observation.' George and Bennett (2005: 114–15) suggest that such deviant cases may also yield information about previously unidentified causal mechanisms that may also operate in other cases (see also Brady and Collier 2004).

Case studies of this type can therefore be theory-confirming or theory-infirming (Lijphart 1971). A theory-confirming case study is one which tends to support the original theory, and shows that it has wider applicability beyond its original context. It thus adds further empirical support to the original theory. By contrast, a theory-infirming case study is one that does not support the original theory, and shows that it is found wanting in some way. However, it is not possible to go as far as to say that a theory-infirming case study can be used to reject the original theory, since just because the theory is found not to apply to a single case, it does not mean that it does not (or will not) apply to other cases.

For this reason, the idea of what has been termed a **crucial case** (Eckstein 1975) is controversial. A case is crucial if it can be regarded as crucial for the confirmation or disconfirmation of a theory. Case selection is based on the likelihood or unlikelihood of the outcome of interest occurring (Bennet and Elman 2006). According to Gerring (2007b: 232)

> a most-likely case is one that, on all dimensions except the dimension of theoretical interest, is predicted to achieve a certain outcome and yet does not. It is therefore disconfirmatory. A least likely case is one that, on all dimensions except the dimension of theoretical interest, is predicted not to achieve a certain outcome and yet does so. It is confirmatory.

The 'least-likely' case study relies on what Levy (2002: 144) has labelled the Sinatra inference: if the theory can make it here, it can make it anywhere. However, despite providing a tough empirical test for the theory in question, it is debatable whether any one-off case can ever be regarded as critical to confirming or disconfirming a theory (see Gerring 2007b for a detailed discussion).

But perhaps the most important use of the case study is to *generate new theory* and hypotheses in areas where no theory exists. The generation of new hypotheses and theory often develops from the analysis of particular case studies. Such case studies can be of great theoretical value, as the examples mentioned earlier attest. They tend to work best when the case selected for analysis provides what Naroll (1966) calls a sort of 'crucial experiment' in which certain variables of interest happen to be present in a special way.

Recently, there has been a great deal of discussion about the merits of using case studies to investigate causal mechanisms underpinning the association between two variables (Brady and Collier 2004; George and Bennett 2005). This is often described as process tracing. It involves looking for evidence of the pressures, incentives, motivations, and decision-making calculus in any given instance of action (George and Bennett 2005), often through the use of elite interview (see Chapter 12). It thus refers to a method of data collection and analysis, and although it is frequently associated with the case study, it can also be used for small-N analysis. Indeed, quantitative analysis often tests for the presence of causal mechanisms as well.

Case studies are widely used in comparative politics. They have a number of key strengths. They allow for a detailed analysis of political phenomena, with rich textual description. Because of this they do not tend to operate at a high level of theoretical abstraction and there is a good match between theory and evidence. They thus generally have relatively high internal validity. However, despite these strengths, they also suffer from a number of fairly obvious limitations, particularly in terms of how far the findings from one case study may be generalized to other contexts. It is not always apparent what lessons, if any, may be learned from a case study until the same type of analysis is repeated elsewhere. Indeed, even when apparently similar hypotheses have been tested in different contexts, it can still be difficult to draw any firm conclusions about whether or not the results confirm (or infirm) the original theory unless the hypotheses have been tested in exactly the same way in each case. Differences in the results may be down to differences in method, model specification (which factors are considered), or measurement (how key concepts are operationalized). One obvious way to reduce some of these problems, then, is simply to increase the number of cases considered in the analysis. This type of comparative research is known as small-N comparison, where N refers to the number of cases that are examined.

Small-N comparison

A lot of comparative research is based on small-N samples. These typically involve the comparison of two or more cases, and although there is no real upper limit to the number of cases that may be examined, it normally does not exceed more than a dozen or so. Small-N comparative studies of this type are widely used in political research, and in some respects are the quintessential form of comparative analysis. Indeed, Lijphart (1971) defines the

comparative approach in terms of small-N analysis (in contrast to the statistical approach of large-N studies). Classic small-N studies include Michael Lewis-Beck's (1986) four-nation study of economic voting in Britain, France, Germany, and Italy; Seymour Martin Lipset's (1959) study of the social requisites of democracy in Europe and South America; Thelda Skocpol's (1979) study of revolution in Russia, France, and China; Barrington Moore's (1966) study of democracy and dictatorship, and Almond and Verba's (1963) study of civic culture in the UK, United States, Germany, Italy, and Mexico. In addition, there are numerous area studies that rely on small-N comparisons to examine a variety of topics, such as ethnic politics in Africa (Posner 2004), vote buying in India (Heath and Tillin 2018), party systems in Eastern Europe (Kitschelt et al. 1999), and democratic stability in Latin America (Mainwaring and Scully 1995), to name but a few.

Comparisons of this type, based on the systematic analysis of several countries, have many advantages. They allow for the detailed in-depth analysis of the case study, but at the same time provide greater scope for contextualization. As such, they leave out neither the particular nor the general. However, although these small-N studies can be extremely useful for generating new ideas and answering questions, they are also fraught with problems, and the careless selection of cases for analysis can lead to the production of very misleading results. When we only analyse a few cases, it makes a great deal of difference which cases we analyse. Indeed, if we wanted to, it would be very easy to fix the results to get almost any answer that we wanted by just handpicking three or four cases that appear to confirm our hypotheses and ignoring all the other cases which do not.

Therefore to make an argument convincing, a number of important considerations need to be taken into account, and the justification for case selection needs to be made clear. If cases are selected in a haphazard way or without care and attention, then there is the very real risk of introducing selection bias. This is something that the researcher needs to be acutely aware of. To a greater or lesser extent problems associated with selection bias affect all small-N comparisons, since we are only dealing with a limited number of cases. We therefore have to pay special attention to how we select cases in order to try and minimize the risk of it occurring.

Small-N comparisons are based on the strategic selection of cases appropriate to test a theory. Which cases we select therefore depends upon the theory we are interested in testing. But at the same time, there are a number of general principles that guide case selection. We are rarely if ever interested in comparing two (or more) cases that are exactly the same in every respect or completely different in every respect. Besides the almost practical impossibility of finding two such countries, even if we could, there would be little analytical point to such an exercise, since the comparison would be redundant (we would basically just be looking at the same thing twice). Rather, we are generally interested in comparing cases that are similar in some respects and different in other respects. We try to take advantage of the similarities and differences we see in the real world to see if the pattern of observed variations is consistent (or not) with our theoretical expectations. In this way, we can test hypotheses that move beyond the purely domestic to examine how the characteristics of countries and the relationship between countries influence political phenomena.

There are two main approaches that have been used to select cases for small-N studies. These are the **Most Similar Systems Design** (MSSD) and **Most Different Systems Design** (MDSD) (Przeworski and Teune 1970). These are based, respectively, on Mills' Method of

Difference and Method of Disagreement. Each approach makes slightly different use of the observed similarities and differences that we observe between countries, though the logic of each is quite similar. We discuss each of these in detail later.

Most Similar Systems Design

The MSSD is based on selecting countries that share many (theoretically) important characteristics, but differ in one crucial respect (related to the hypothesis of interest). This sort of approach is frequently used in area studies (Przeworski and Teune 1970: 33), where countries within a specific region, such as Africa (or Latin America, East Europe, or South Asia, etc.), are similar in many important respects, perhaps to do with their culture, history, level of economic development, and social structure. The shared characteristics act as a control in order to test whether the crucial difference between the countries is associated with the variation in the dependent variable.

Table 9.1 provides a conceptual overview of this method. The columns refer to the selected cases and the rows refer to the variables of interest. We can see that the selected cases are similar with respect to variables A, B, and C. These variables constitute the control, and cannot meaningfully explain the variation in the dependent variable since they are the same for all cases. However, importantly, the selected cases do vary on variable X. This is the variable related to our key hypothesis, which we think helps to explain the variation in our dependent variable Y. If our hypothesis is supported, then we should observe that our dependent variable Y varies in accordance with the variation in X. Thus the presence (or absence) of the key explanatory factor is able to account for the variation in the outcome of interest.

Example 1: the political salience of ethnic cleavages

We can illustrate the logic of this selection method with an example. A classic study of this type was carried out by Daniel Posner (2004) on the political salience of ethnic cleavages. Posner wanted to explore the factors that are associated with the political salience of cultural difference, and in particular, the factors associated with whether different ethnic groups were political allies or rivals. His hypothesis is that the political salience of ethnic difference is related to the population size of the groups in question. When the ethnic groups are both large, they constitute a viable base for political support, and so politicians

Table 9.1 Most Similar Systems Design

Case 1	Case 2	Case n	
A	A	A	Overall similarities
B	B	B	
C	C	C	
X	Not X	Not X	Crucial differences
Y	Not Y	Not Y	

will mobilize the groups, and the cleavage that divides them will become politically salient. If, however, the groups are both small, then they will go unmobilized and the cleavage that separates them will remain politically irrelevant. The cultural differences between the groups will still exist, but there will be no political importance attached to them.

Using an MSSD, Posner examined the political salience of ethnic differences between two tribes, the Chewas and Tumbukas, who constituted a large proportion of the electorate in Malawi, but only a small proportion of the electorate in Zambia (Table 9.2). In doing so, Posner was able to control for a wide variety of factors that previous research suggested was important. The key thing to remember is that the selection of cases is based, in theory, on the independent variables, not on the dependent variable. The control variables are the factors that the ethnic groups in each country share. Ideally, our control variables should include variables that are theoretically important, since failure to include an important factor may result in omitted **variable bias**. Omitted variable bias, as the name suggests, occurs when we fail to take into account important explanatory factors. The consequence of this is that factors that are not really causally important may appear important because they are associated with (or have even influenced) the unmeasured missing variable. This can cause **spurious association** (see Chapter 17). The extent to which we can have confidence that our results will stand up and be robust is therefore dependent in part upon how adequately we have managed to control for other potentially important factors.

Posner's study controls for four important factors, variables A to D, which are all thought to increase the salience of ethnic differences. Variable A refers to the racial differences between the two groups, variable B refers to the cultural differences between the two groups, variable C refers to the historical development of the political system, and variable D refers to the experience of colonialism. Since these factors are the same for each of the countries, they act as a control. If we control for theoretically important variables, and our variable of interest appears to matter, then we can have greater confidence in the results. Importantly, then, the variable that relates to the hypothesis we are interested in testing varies across our cases. In Posner's example, variable X refers to the size of the ethnic groups.

Following Posner, our expectation is that the variation in the size of the ethnic groups (our explanatory factor) will be associated with variation in the salience of ethnic cleavages (the

Table 9.2 Most Similar Systems Design: the political salience of ethnic differences

Zambia	Malawi	
A	A	Overall similarities
B	B	
C	C	
D	D	
E = 0	E = 1	Crucial differences
Low political salience	High political salience	

Notes: A = racial differences between groups;

B = cultural differences between groups;

C = historical development of political system;

D = experience of colonialism;

E = (0 = ethnic groups are small) (1 = ethnic groups are large).

thing we are trying to explain). That is, when the ethnic groups are both large, the cleavage that divides them will be politically salient and so they will be rivals, but when the groups are both small, the cleavage that separates them will remain politically irrelevant and so they will be allies. If we observe this pattern, then our hypothesis is supported. If we do not, then our hypothesis is not supported. From Table 9.2 we can see that the hypothesis does indeed receive empirical support, and that the political salience of ethnic differences is higher in Malawi, where the ethnic groups are both large, than it is in Zambia, where the groups are both small.

Of course, when we are dealing with only a few cases, there are always going to be lots of potentially important differences between the cases, and it is impossible to completely control for everything that might be theoretically relevant. In this sense, as Przeworski and Teune (1970) suggest, it is easy to overdetermine the dependent variable. Lijphart (1971) describes this as the 'too many variables, too few countries' problem. Whichever cases we select there will always be a number of differences between them, which makes it very difficult to establish which are the crucial differences and which are not.

For example, suppose we were interested in explaining the sources of democratic stability. Our hypothesis might be that democratic stability is related to the institutionalization of political parties in society. That is, where political parties have deep roots and are embedded within society, democracy tends to be more stable. To test this hypothesis, we might compare India and Pakistan, since they share many important similarities. They shared the experience of colonial rule under the British, became independent at the same time, and were both poor with low literacy rates and a largely agricultural economy at the time of democratization. In many respects, then, they were similar on theoretically important independent variables. These shared characteristics are unable to explain any differences between the two countries, such as why democracy has persisted in India over the past fifty years, but struggled in Pakistan, with frequent episodes of authoritarian rule. Our key independent variable of party system institutionalization does, however, appear to vary in the expected way between the two countries. Whereas the Congress Party in India had a strong base right across the country, penetrating even the most remote rural villages, the Muslim League in Pakistan was much more weakly institutionalized.

However, there are also many potentially relevant differences between the countries, to do with leadership (Nehru versus Jinnah), social demography and religion, institutional capacity, international support, and so on. It is difficult to establish, using pair-wise comparison, which of these differences is the crucial causal difference, although some candidates will be more plausible than others. It is also difficult to establish what combination of these variables is important. If we were to add more cases to our analysis, we might be able to eliminate some of the spurious differences that we observe between the two countries, but in reality there are rarely enough cases to find the right combination of similarities and differences we need to exactly test our theory.

Most Different Systems Design

The logic of MDSD is to select cases that are different in most respects and only similar on the key explanatory variable of interest. Thus, rather than comparing countries that are very similar to each other, we compare countries that are very different from each other, hence the name. With this approach, the observed differences on important variables act as

Table 9.3 Most Different Systems Design

Case 1	Case 2	Case n	
A	Not A	Not A	
B	Not B	Not B	Overall differences
C	Not C	Not C	
X	X	X	Crucial similarity
Y	Y	Y	

a control in order to test whether the crucial similarity that the countries share is associated with the dependent variable.

Table 9.3 provides a conceptual overview of this method. Once again, the columns refer to the selected cases and the rows refer to the variables in the analysis. This time we can see that the cases are different with respect to variables A, B, and C. These variables constitute the control. Importantly, our cases are all similar with respect to the key independent variable of interest, X. If our hypothesis is supported, then we should observe that our dependent variable Y is also similar across our cases. Thus the presence of the key explanatory factor is associated with the presence of the outcome of interest.

Example 2: revolutions

We can illustrate this approach with the example of Theda Skocpol's (1979) famous study of revolutions, which was based in part upon an MDSD, comparing the causes of social revolution in three very different countries, Russia (1917–21), France (1787–1800), and China (1911–49) (Table 9.4). (See Geddes 2003 for an extended discussion of this example.) Broadly speaking, Skocpol's main argument is that social revolutions occur when external military threats provoke a split in the ruling elite and peasant communities take advantage of this split and revolt. Accordingly, countries which experience state breakdown and peasant revolt will also tend to experience social revolutions, even if they differ in many other important respects.

Table 9.4 Most Different Systems Design: the causes of revolution

France	Russia	China	
A	Not A	Not A	
B	Not B	Not B	Overall differences
C	Not C	Not C	
X-Elite split	X-Elite split	X-Elite split	Crucial similarity
Y-Revolution	Y-Revolution	Y-Revolution	

Notes: A = political system;

B = social and cultural context;

C = economic conditions.

In order to test this hypothesis, Skocpol selected three countries very different in social, economic, and political terms, but similar in terms of having autonomous peasant populations, and similar in that they all experienced periods of external military threats which produced splits in their ruling elite. The key thing to remember with classic MDSDs is that the selection of cases should be based on the independent variables, not on the dependent variable. In the case of Skocpol's study, it is debatable whether this decision rule was followed (we will return to this later). In this example, then, the control variables are the factors that are different for each country, such as their social, economic, and political backgrounds. The key independent variables of interest are the same for each country: they all had elite splits, peasant uprisings, and state collapse; and, crucially, the outcome variable is the same for each country as well—they all experienced social revolutions, thus confirming the initial hypothesis.

Issues in small-N research

The issue of case selection in small-N studies is extremely important because the cases we choose to analyse can influence the answers we get to a particular question, and this is particularly true when we are only analysing a few cases. When this occurs, it is known as selection bias. If we do not think carefully about case selection, we can end up with answers that are not very robust or even somewhat misleading. That is, we will make faulty inferences. The best way to avoid this happening is to conduct a large-N study and select cases using random probability methods or to take a full census of cases. But this is not always a viable option. To a certain extent, then, selection bias, or at least the risk of selection bias, is always going to be a danger in small-N studies, particularly when the study is designed to try and shed light on a broader phenomenon, and make some sort of generalization, however tentatively put.

Discussion about the relative merits of different strategies for selecting cases in small-N research has been among the most controversial and hotly debated issues in comparative research (for different sides of the debate, see King, Keohane, and Verba 1994; Ragin 2000, 2008; Geddes 2003; Collier, Mahoney, and Seawright 2004; George and Bennett 2005). One of the main strategies that has been proposed is to select cases according to their characteristics on the independent variables (see King, Keohane, and Verba 1994; Geddes 2003). This strategy has been elevated to almost law-like status: never select on the dependent variable. Some have even gone as far as to say that selecting on the dependent variable is a type of 'inferential felony' (Achen and Snidal 1989). This position has been challenged by some scholars (see Collier and Mahoney 1996; Dion 1998), but before outlining the criticisms, which we will return to later, it is worth discussing the logic behind the initial decision rule.

The argument goes that if we select on the dependent variable and do not pay sufficient attention to the independent variables, we may unwittingly exclude important factors from our analysis, or fail to control for them adequately. This means that what looks like a strong causal relationship between two variables may in fact be spurious. For example, suppose we want to examine the representation of women in parliament. We might choose three countries which have a relatively high percentage of women in parliament, such as Sweden (45 per cent), Norway (40 per cent), and Finland (40 per cent), and seek to identify

what factors they have in common. Since all three cases have electoral systems based on proportional representation, we might conclude that this is an important factor, and infer that proportional representation systems are conducive to promoting the representation of women in parliament. But this inference would be wrong. There are many other countries with proportional representation systems that do not have a high percentage of women in parliament, such as Brazil (8 per cent), Turkey (9 per cent), and Uruguay (15 per cent). The percentage of women in parliament in Scandinavia may therefore have little to do with the electoral system, and more to do with the strategy of political recruitment by political elites or the political culture of the general public, both of which may be exogenous to institutional design. However, because we have not taken these factors into account when selecting our cases, it is difficult to establish which, if any, are important. The problem with not selecting cases on the independent variable, then, is that it makes it very difficult to control for potentially important variables.

When research is theory-driven, or at least research-guided (see Chapters 4, 5, and 6), we should always have some hunch that guides our research, and accordingly we should design our research in order to provide an appropriate 'test' of this hunch. Therefore, if our hunch is that the type of electoral system is an important factor in determining the representation of women in parliament, we would want to select cases that allow us to explore this, which means selecting cases according to the independent variables. In small-N analysis, this can be done either by selecting cases that are very similar to each other but vary on the key independent variable (MSSD) or selecting cases that are very different to each other but are similar on the key independent variable (MDSD). Both these approaches allow us to control for potentially important factors, and so reduce the risk of obtaining spurious relationships.

When carrying out MDSD, it is easy to fall into the trap of selecting cases on the dependent variable. Indeed, MDSD is often incorrectly characterized as being based on this strategy. For example, if we were interested in exploring the cause of revolutions, we might be tempted to select three countries which have experienced revolutions and see what they have in common. This is, in essence, what Skocpol did in her study of revolutions, which has been criticized by Geddes (2003). Although Skocpol's study has been incredibly influential, it is also widely regarded as being based on spurious inference, in so far as the main causes of revolution do not appear to be generalizable to other relevant cases (see Geddes 2003). Geddes argues that there are lots of cases that do not conform to Skocpol's theoretical expectations, where there have been instances of external threat which did not lead to a social revolution. Although in some ways this issue is part and parcel of small-N research, there are also some specific issues to do with case selection that may have contributed towards this problem in Skocpol's study. Geddes argues that, in essence, what Skocpol did was select her cases on the dependent variable. That is, she didn't select them because they had all experienced peasant uprisings and state collapse, but because they had all experienced a social revolution.

However, selection on the dependent variable can just as easily occur in Most Similar System Designs. For example, it would not be wise to select two cases that have populist leaders and two cases that do not and look to see how they are different, because if you do, you are just as likely to introduce the same kind of selection errors. In both cases, you risk leaving out or not properly considering important explanatory factors, which will bias the

results. The best way of ensuring that this doesn't happen (or at least is less likely to happen) is to base the selection on the independent variables. If prior theory suggests, for example, that institutional weakness is an important source of populism, then it is a good idea to take this into account when you select your cases. Thus, whether you are carrying out MSSD or MDSD, in theory the values of the cases on the dependent variable should not be a criterion for selection, and all selection decisions should be based on the values of the cases on the independent variables only. Of course, in practice, we often cannot help but know what value the dependent variable takes for each of our cases, but the important thing is that this knowledge should not influence our selection decisions (and if it does, we risk introducing greater selection error).

With exploratory research, there is sometimes a justification for selecting on the dependent variable if the study is in a genuinely new field and prior research does not suggest any important explanatory factors. Under these rare circumstances, it is not possible to control for theoretically important independent variables because no such background literature exists. However, instances of new research areas that have absolutely no connection to previous research are exceptionally rare. And even where they do exist, the researcher generally still poses some guiding themes or foreshadowed problems (see Chapter 13) that they want to explore, and so these factors should also feed into the selection of cases.

In recent years, there has been a bit of a qualitative backlash against this principle of selection, perhaps partly in response to the overtly statistical language in which King, Keohane, and Verba (1994) originally framed their decision rule (see Collier and Mahoney 1996; Ragin 2000, 2008; Collier, Mahoney, and Seawright 2004: 85–102; George and Bennett 2005). In essence, what these qualitative methodologists argue is that when certain qualitative methodologies are employed, particularly process tracing, it is permissible to select on the dependent variable. According to George and Bennett (2005), process tracing is fundamentally different from methods that rely on 'co-variation', and the method's contribution to causal inference arises from its evidence on the process which connects the cause and the outcome. Accordingly, to scholars like Collier, Mahoney, and Seawright (2004) and Bennett and Elman (2006), it is therefore not susceptible to selection bias, which means that the principles of case selection discussed earlier are largely irrelevant, and that cases may be selected on the basis of the dependent variable (see Bennett and Elman 2006 for further discussion on this point). This is fine up to a point, at least in terms of describing what happens within a given case (internal validity). But as soon as we are interested in comparison and making wider inferences (external validity), it is a good idea to take the issue of case selection seriously, or else we can easily end up with misleading results (see Chapter 11 on sampling principles).

A second criticism that is often made against the principle of never selecting on the dependent variable is that there is a distinction between necessary and sufficient causes. Accordingly, selecting on the dependent variable, and looking to see what the countries have in common, can reveal the necessary conditions (or causes) for an outcome of interest, even if these causes are not in themselves sufficient to produce or cause the outcome of interest (see Dion 1998; Braumoeller and Goertz 2000; and Goertz and Starr 2003). For example, with reference to Skocpol, Dion (1998) distinguishes between a claim based on a sufficient condition for something to happen, such as 'state crisis leads to social revolution', and a claim based on a necessary condition for something to happen, such as 'social revolutions

arise only if there is a state crisis'. Whereas state crisis which does not lead to social revolution is evidence against the sufficient proposition, only a social revolution which did not experience state crisis is evidence against the necessary condition. Therefore, in order to test hypotheses based on necessary causes, it is permissible to select countries which all have the same outcome on the dependent variable and see what they have in common.

Although this is a valid point, it might be argued that it is often useful to know whether the condition is only necessary or whether it is necessary and sufficient, in which case selecting on the dependent variable is not of much help. It might also be argued that it is difficult to distinguish between a necessary condition and a spurious condition. For example, suppose we are interested in democracy, and want to investigate the claim that 'democracy only persists over the long term in mono-ethnic societies'. If we select on the dependent variable and only choose countries which have been democratic for more than twenty years, the chances are that they will be ethnically homogeneous, confirming our hypothesis. But this does not mean that ethnic homogeneity is a necessary cause for democratic persistence. It is in fact not a cause at all, and the apparent coincidence between democratic stability and ethnic homogeneity is spurious, driven by the association between ethnic diversity and economic development (see Chapter 17). The only way we can try and knock down spurious relationships is to take seriously the idea of controlling for theoretically important factors, and the most effective way in which we can do this is with reference to selecting cases on the basis of the independent variables.

This brings us to a second related point, which is often confused with the principle of not selecting on the dependent variable. Many people believe that it is a bad idea not to have any variation in the dependent variable. That is, it is desirable to have negative cases as well as positive cases, such as cases where there has not been ethnic conflict, as well as cases where there has been ethnic conflict. This suggests that an MSSD is preferable to an MDSD, since the latter is often unfairly criticized as selecting countries with the same outcome variables. But the logic underlying this supposition is faulty. Selecting cases so that there is variation in the dependent variable is still selecting cases on the dependent variable. The problems of selection bias do not come from the fact that there is no variation in the dependent variable, but from the fact that important independent variables have not been properly controlled for.

The key then is to control for theoretically important independent variables in order to test a specific hypothesis. Whether the approach used is MDSD or MSSD, failing to control for theoretically important variables can introduce error into the analysis and risk producing unreliable results. The risk of this occurring is no greater when case selection is associated with similarities in the outcome variable than when it is associated with differences in the outcome variable. However, these problems can be compounded. The worst-case scenario is to select on the dependent variable and choose similar types of countries with a similar outcome variable; or conversely to select different types of countries with a different outcome variable. The former is perhaps more common in comparative research and is a particular problem in area studies, where the dependent variable is the same for all selected cases. For example, studies on the rise of the 'New Left' in Latin America may seek to identify the common factors underpinning the electoral success of leftist leaders like Hugo Chávez in Venezuela, Evo Morales in Bolivia, and Rafael Correa in Ecuador. Given that these countries share many similarities, and that many of these similarities may also

be shared with other countries in the region which have not experienced the sudden rise of 'New Left' leaders, it would be easy to ascribe causal significance to similarities between the three countries that are not really that important at all.

When we are dealing with just a small number of cases, selection bias is always likely to be a problem. We cannot get rid of it completely, and the harsh reality is that we cannot aim to make robust generalizations when we have a sample of just two or three cases. For this reason small-N studies are generally better at generating rather than testing theory. It is difficult ever to say with confidence that a theory is good or bad when it is just tested against a small number of cases. It can apply or not, but until it has been applied to a large number of cases, it is difficult to draw any strong conclusions. The only way we can have confidence that our findings can be robustly generalized and are not based on spurious associations is to gradually extend the analysis and test the key hypotheses across more and more cases. This can be done either by replicating small-N studies using different cases, or by carrying out large-N quantitative research that seeks to test more widely insights generated from small-N analysis.

Of course, under many circumstances, small-N research is the only option. Often the population of interest is not the whole world, but only a handful of countries. For example, if we are interested in the success (or failure) of market reforms in fragile democracies, then there is only a limited number of countries that would be eligible for inclusion (Weyland 2002). But this does not change the basic point. Until the theory has been tested on a large number of cases (whether they exist or not), then the wider implications of the findings must be treated with caution.

The comparative inferences we can draw from small-N studies are therefore much more tentative than the inferences we can draw from large-N studies. This is because the standard for finding supporting evidence in favour (or against) a hypothesis tends to be somewhat different. In quantitative analysis (as we discuss in Chapters 15 to 17), the presence (or absence) of causal relationships are defined probabilistically. That is, for a hypothesis to be supported (or rejected), we make a decision based on the balance of probability. However, in small-N studies, it becomes more difficult to talk about probabilistic causality. If we compare three cases and observe the expected relationship in two of them, do we reject or accept our working hypothesis? Generally, in small-N studies, causality is deterministic. That is, in order to find evidence to support a hypothesis, the expected pattern needs to be observed in all the cases. This obviously creates a problem in that it is quite easy to reject a hypothesis that is actually true, just because one of the cases deviates from the expected pattern.

Qualitative comparative analysis (intermediate N)

Recently, discussion of small-N comparison has moved in a number of different directions with the development of Qualitative Comparative Analysis (QCA). QCA draws on insights from the methodology of necessary and sufficient conditions, Boolean algebra, and fuzzy-set logic (e.g. Ragin 1987, 2000, 2008; Goertz and Starr 2003; Hacket 2015, 2016). QCA has grown in popularity over the past few years, and there are now a whole host of textbooks devoted to explaining the method (see Ragin 1987, 2000, 2008; Caramani 2009; Rihoux and Ragin 2009; http://www.compasss.org). Typically, qualitative researchers examine a few

cases at a time in detail, and look at how the different parts of a case fit together, both contextually and historically. According to Ragin, by formalizing the logic of qualitative analysis, QCA makes it possible to bring the logic and empirical intensity of qualitative approaches to studies that embrace more than a handful of cases, which would normally be analysed using variable-orientated, quantitative methods. In some ways, then, QCA occupies an intermediary position between small-N analysis and large-N analysis, and is generally used for samples that range from about a dozen or so cases to several hundred.

Boolean methods of logical comparison represent each case as a combination of causal and outcome conditions. These combinations can be compared with each other and then logically simplified through a process of paired comparison. Computer algorithms can be used to provide techniques for simplifying this type of data (software is available on Charles Ragin's website at: http://www.u.arizona.edu/~cragin/fsQCA). But when only a few cases are used, it is also possible to do it by hand. The data matrix is reformulated as a **truth table** (see Ragin 1987).

Broadly speaking, there are two main ways of doing QCA. The first is known as **crisp set** QCA, in which all the variables in the analysis are treated as simple dichotomies (i.e. the hypothesized causal condition is either present or absent). The second is known as **fuzzy set** QCA, in which variables are allowed to take different values and are calibrated on an interval scale between 0.0 and 1.0 (see Ragin 2008: chapter 2). A conventional (or crisp) set is dichotomous: a case is either 'in' or 'out' of a set, for example, the set of democracies. Thus, a conventional set is comparable to a binary variable with two values, 1 ('in', i.e. democracy) and 0 ('out', i.e. non-democracy). A fuzzy set, by contrast, permits membership in the interval between 0 and 1, while retaining the two qualitative states of full membership and full non-membership. Thus, the fuzzy set of democracies could include countries which are 'fully in' the set (fuzzy membership = 1.0), some who are 'almost fully in' the set (membership = 0.90), some who are neither 'more in' nor 'more out' of the set (membership = 0.5; also known as the 'crossover point'), some who are 'barely more out than in' the set (membership = 0.45), and so on, down to those who are 'fully out' of the set (membership = 0). It is up to the researcher to specify procedures for assigning fuzzy membership scores to cases, and these procedures must be both open and explicit so that they can be evaluated by other scholars (see Ragin 2008). Despite these measurement differences, the general analytical principles of the two approaches are much the same. For the sake of simplicity, we will focus just on crisp sets to illustrate how the approach can work.

QCA is a powerful heuristic tool as it can be used for several purposes, such as summarizing data, producing typologies, and elaborating new theories or models, as well as for testing existing theories, models, and hypotheses (Rihoux 2006). In particular, QCA is well-suited to unravelling causal complexity in order to detect the different conditions (or configurations) that can lead to the same outcome occurring (which is sometimes called equifinality or multiple causation). The way in which it does this is by utilizing what are known as truth tables. Truth tables present all the logical possible combinations of conditions and the outcome associated with each combination. By examining all the possible different combinations of variables that are associated with an outcome occurring (or not), the main aim of the approach is to try and simplify these different combinations to as few distinct (necessary) combinations as possible. This is done using something called Boolean algebra (see Ragin 2008 for a step-by-step guide to how this works in practice).

We can illustrate how this approach works by drawing on a hypothetical example. Table 9.5 is based on Ragin's (1987: 96) hypothetical truth table on the factors associated with the success of strike activity. In this example, the aim of the research is to explore how three main factors influence whether or not a strike is successful. The three causal conditions are A = booming product market; B = threat of sympathy strikes; and C = large strike fund. Data are collected on thirty-nine strikes, and for each case the causal conditions are (hypothetically) coded 1 if they are thought to be present (and 0 otherwise). Similarly, the outcome variable is coded 1 if the strike was successful, and 0 if it was not. The first three columns of the truth table refer to the independent variables/causal conditions. The rows refer to all the logical combinations that these conditions can take. When there are three independent variables, there are eight possible combinations in which they can be arranged. In more complex analysis, the rows (representing combinations of causal conditions) may be quite numerous. The fourth column refers to the dependent variable, and shows the recorded outcome (whether or not the strike was successful) for each combination of conditions. The last column refers to the number of cases that were observed in each row (though these are not really used in the subsequent analysis).

From Table 9.5 we can see that the four different combinations of conditions all led to a successful strike (where S = 1). We can express these conditions in an equation as follows:

$$S = ABC + AbC + ABc + aBc$$

where a capital letter indicates that the causal condition in question was present and a lower-case letter indicates that it was absent.

The first step in the Boolean analysis of these data is to try and combine as many compatible rows of the truth table as possible. In a sense, this involves eliminating all redundant terms. If two combinations differ in only one factor but produce the same result, then the factor they differ on can be thought of as irrelevant.

Table 9.5 Hypothetical truth table showing three causes of successful strikes

Condition			Outcome	Frequency
A	B	C	S	N
0	0	0	0	4
0	0	1	0	6
0	1	0	1	5
0	1	1	0	3
1	0	0	0	9
1	0	1	1	6
1	1	0	1	2
1	1	1	1	4

Notes: A = booming product market;
B = threat of sympathy strikes;
C = large strike fund;
S = success of strike;
N = number of cases.

So, for example, since ABC and AbC are both associated with the presence of outcome S = 1, we may suspect that B is superfluous. That is, when we have A and C together, it doesn't matter whether we have B or not in order for S to occur. We can therefore combine the first two terms in the equation to produce AC.

Similarly, when B is present and C is not present (Bc), it doesn't appear to make any difference whether or not A is present in order for S to occur. So we can combine the terms ABc and aBc to produce Bc.

We can therefore express the equation in slightly simplified terms as:

S = AC + Bc

This final equation states simply that successful strikes occur when there is a booming market for the product produced by the workers *and* a large available strike fund (AC) or when there is the threat of sympathy strikes by workers in associated industries combined with a low strike fund (Bc). This finding suggests that strikes can be successful for different reasons. It also suggests that whereas the impact of A and C is conditional on the presence of each other, the impact of B is conditional on the absence of C. That is, the threat of sympathy strikes is taken seriously only when the striking workers badly need the support of other workers.

In order to use the truth table presented in Table 9.5, it is necessary to determine an output value for each row that is either 1 or 0. With hypothetical data, this is not a problem, but the real world is not so neat and the use of truth tables is often complicated by the presence of contradictory outcomes. That is where cases which share the same causal conditions exhibit different outcomes on the variable of interest. It is therefore more difficult to illustrate how this method works with real data.

Nonetheless, this type of approach has been used, adapted, and developed for a variety of purposes and to explore a variety of different political phenomena. One common application of the approach is to examine the extent to which different pathways (or combinations of causal conditions) can lead to similar outcomes. Recent examples include Rihoux's (2006) analysis of the success of Green parties in Western Europe; Veugelers and Magnan's (2005) analysis of the success of Far Right parties in Western Europe; and Berg-Schlosser's (2007) analysis of the success of democracy in Africa. Berg-Schlosser (2007) uses QCA to identify the different pathways to democratic success in Africa. He identifies four main groups of countries, in which each group developed democracy under somewhat different conditions. For each group, 'loser acceptance' at the time of transition turned out to be a necessary condition for the long-term success of democracy, but by itself this factor was not sufficient for democracy to prevail, as is demonstrated in the negative examples of Congo-Brazzaville, Gambia, and others where democracy was not sustained.

Although QCA provides an interesting approach for examining causal complexity, it is also beset with many of the same problems that we discussed earlier with reference to case selection in small-N research. There are a number of issues we need to pay attention to when carrying out QCA. The first issue refers to how many cases we choose to examine in our analysis. In order to examine whether the impact of a particular factor is necessary or sufficient, and whether its impact is contingent on the presence or absence of other variables, it is desirable to have cases which capture the different possible combinations of causal conditions, so, wherever possible, outcomes are known for each row of the truth

table. Of course, in reality, some combinations are simply not present in the real world. Nonetheless, when very small samples are used, there are going to be a lot of gaps in the truth table.

Second, there is also the issue of case selection. As with all comparative research, it is necessary to define the theoretical population of interest from which the cases are selected. Sometimes all the countries (or cases) in a given population can be examined, but at other times a sample will be needed. As soon as we are dealing with a sample, we need to think about case selection or else our results may end up being very misleading. However, even when we are analysing an entire population of cases, we should still think about how robust our findings are. When we use exploratory techniques such as QCA to mine the data for patterns, we have to be sensitive to the possibility that our findings are driven by quirks or anomalies in the data that might not be present if we were to try and reproduce our analysis with different (but similar) cases. For this reason, scholars who use QCA often go on to use other methods of inquiry to test their arguments more widely, and so will think about the ways in which QCA can be used to complement large-N research (see Paine 2016; Thiem and Baumgartner 2016).

Quantitative analysis (large-N comparison)

Large-N comparative studies are widely used in political research. Studies of this type frequently employ quantitative analysis to answer questions on topics as diverse as the nature and direction of the relationship between economic development and democracy (Przeworski et al. 2000); the causes of ethnic conflict (Wilkinson 2004) and civil war (Gleditsch 2007); the impact of party competition on social welfare (Lipsmeyer 2002), and the impact of international intervention on civil war (Cunningham 2016) to name but a few.

The principles of inference and quantitative data analysis are discussed elsewhere in this book (see in particular Chapters 16 and 17) and so will not be dealt with here. Instead, we will briefly discuss issues to do with selecting cases and data collection. Large-N comparative research, based on the statistical analysis of many countries, has many advantages. It allows us to rigorously test different hypotheses and make inferences about how variables are connected. Many of the problems of selection bias associated with small-N analysis that we discussed earlier can therefore be mitigated by using this approach. Large-N comparisons also allow us to systematically examine the effect of many different variables and to knock down spurious associations. They also allow us to examine how different factors interact with each other to produce different consequences in different contexts.

This allows us to have greater confidence that our results are robust. This confidence comes in part from simply being able to analyse a large number of cases (which are broadly representative of the population we are studying) and in part from being able to examine the impact of many different variables simultaneously, so as to rule out plausible alternative explanations. To rule out plausible alternatives we need to analyse many cases, or else we may encounter the 'too many variables not enough cases problem' that we discussed earlier. However, despite these obvious strengths, analysing a lot of cases also brings its own problems. This is often presented in terms of a trade-off between internal and external validity. As Coppedge (1999: 464) puts it, although large-N quantitative analysis is the

best method available for testing generalizations, especially generalizations about complex causal relationships, it is also often criticized for its 'thin' (reductionist or simplistic) concepts and theories.

Large-N comparative studies involve the tasks of case selection, data collection, and data analysis. The principle of case selection for large-N research is fairly clear-cut, and there is a generally agreed upon set of procedures of best practice. Ultimately, the goal of case selection is to achieve a representative sample of the **population** of interest which is of sufficient sample size to enable robust inferences to be made. The principles of sample design are discussed in detail in Chapter 11, and many of the same principles apply to case selection in comparative research. The main advantage of this approach is that it helps us to overcome many of the problems to do with selection bias that we discussed earlier in the chapter.

To test any theory or hypothesis one must first identify the domain of the argument or the universe of cases to which the hypothesis should apply. This is generally referred to as the population of interest. So, for example, if you are interested in exploring why some countries engage in war whereas others don't, the population of interest may be all the countries in the world, since every country either engages in war or doesn't. However, if you're interested in exploring why some countries engage in war in the face of provocation, whereas others don't, your population of interest may be all the countries in the world that have experienced provocation. (You would need to think very carefully about how you could define provocation and then how you could measure it.)

Having defined the population of interest, the next step is to select eligible cases for the study. If the population is very big, it may be necessary to draw a random sample (see Chapter 11). However, it is quite common in large-N comparative research to simply analyse all the cases in the population of interest, and there are even a number of studies which have created data sets for all the countries in the world. The problem with this though is that this is still rather a small sample. These studies generally have data on about 192 countries, and if we had a sample of just 192 people in a survey we would be a bit reluctant to draw any strong conclusions. The same applies to comparative research. The size of the sample in relation to the population doesn't make any difference to how reliable the findings from our sample are. When you have a small sample there is always the risk of drawing faulty inferences.

Comparatavists often try to get round this problem by looking at data over time in order to try and increase their sample size. So, for example, they may collect information on 192 countries at yearly intervals over say a fifty-year period, which would give them a total sample of 9,600. But even quite modest increases in sample size can help to knock down many faulty inferences. Geddes (2003: 89–105) provides a particularly striking example of this, which both highlights the limitations of small-N research and the advantages of large-N research. Geddes recounts how analysts trying to explain why some developing countries have grown economically faster than others have tended to focus their analysis on a few successful new industrializing countries (NICs), such as Taiwan, South Korea, Singapore, Brazil, and Mexico. It was noted that these countries all had fairly repressed labour forces. Accordingly, some scholars asserted that repressive labour laws and weak unions constitute a comparative advantage in international economic competition. The problem with this sort of claim is that it makes sweeping generalizations from the basis of a very few cases. The hypothesis sounds as if it might be plausible. And as we discussed earlier,

small-N research is very useful for developing theories. But to actually test this theory and see whether it really does stand up, we need to see whether it applies more generally. To do this Geddes collected data on economic growth and labour repression for eighty-four developing countries and examined whether the finding still held.

The statistical techniques she used to do this are discussed in Chapters 16 and 17. However, the main point of interest is that she found that the theory did not stand up to wider generalization. Although it was true that Taiwan, South Korea, and the others did have repressive labour laws and high levels of economic growth, Geddes found that there were many other countries that also had repressive labour regulations, such as Iran and Argentina, but which had not experienced economic growth. Similarly, she also found that there were many countries which did not have repressive labour laws, such as Botswana, but had experienced high growth. In short, the initial assumption that labour repression influences economic growth was a faulty inference. It is easy to make these sorts of faulty inferences when we only examine a few cases. And the best way to overcome this problem is to examine many cases. If we are interested in testing theories rigorously, and seeing whether they hold up in the different contexts to which they are thought to apply, then we are inevitably drawn to large-N analysis.

Whereas almost any type of data can be used in small-N research, large-N research generally relies upon quantitative data. We call it quantitative data because all the information is assigned a numerical value. Sometimes how we assign this value is obvious, because the variable of interest is measured in numbers, such as votes or inflation, or GDP per capita, war casualties, or unemployment. At other times, how we assign a value is less obvious, and is the result of a coding decision based on a qualitative judgement to do with how democratic an 'expert' thinks a particular country is. There are many different sources of data that are used in large-N comparative research. Surveys are a valuable resource (discussed in Chapter 11), and there are now many cross-national surveys that have been carried out in almost identical fashion in many different countries around the world over a relatively long period of time. These surveys provide valuable information about how public perceptions of corruption, or support for democracy, or support for human rights vary between different regions. Some of the most widely used cross-national surveys of this type are the Comparative Study of Electoral Systems (CSES), the World Values Survey, Eurobarometer, Latinobarometer, and Afrobarometer.

But population surveys are not the only source of data that large-N studies can draw on. There are now many cross-national data sets that have been compiled from expert surveys (see Chapter 12), which measure perceptions about the level of corruption in a country (Transparency International) or the level of democracy in a country (which forms part of the Freedom House and Polity IV measures of democracy). Comparative research also draws on policy documents and speeches (see Chapter 14) to examine the ideology of governments and political parties (Budge et al. 2001; Benoit and Laver 2006) and leadership styles and populism (Hawkins 2009). These methods of data collection are discussed in detail elsewhere in the book, and so will not be considered here.

But whatever data is collected, it is important that it is both valid and reliable. Broadly speaking, validity refers to whether the data measures what it is supposed to measure. That is, how closely does the operational indicator match the concept of interest? For example, if you are interested in the level of corruption in a country, is a survey question on a business

person's perceptions of corruption a valid measure of this? The issue is not whether you have measured accurately or not whether the business person thinks that the country in question is corrupt or not, but whether perceptions of corruption are a valid measure of actual corruption. Reliability refers to how well the operational indicator has been measured. That is, does the question really tap into what the business person thinks about the country in question?

The principles of data collection in large-N research are that information about the world can be coded or summarized in ways that capture meaning. We start with concepts or constructs that we want to measure; we develop operational indicators for these constructs; and then we collect data to measure the operational indicators (see Chapter 7). Some constructs are fairly easy to measure. Suppose we are interested in conventional political participation. Our conceptual definition of political participation might be whether or not people vote in general elections. Our operational indicator—which is what we use to measure our conceptual definition, might therefore be official election returns from around the world which detail what proportion of the voting-age public actually cast a ballot. This is a pretty valid measure. The data clearly taps into our concept of participation. It is also likely to be a reasonably reliable measure, though practices of vote rigging may make it less reliable in some countries than others.

However, other constructs are more difficult to measure. Suppose you are interested in measuring a country's level of democracy. We must therefore provide a conceptual definition of what we mean by democracy. There is a vast literature on this, and scholars disagree about how democracy should be defined (see Coppedge 2012). Some authors favour a thin definition based on institutional arrangements, others favour a thicker or fuller definition based on civil and political rights. Suppose we go for a thin definition of democracy, in which we define democracy as a system of government in which governments are selected though competitive elections. Our operational definition may therefore be the competitiveness of elections, and our operational indicator of competitiveness might be the difference in vote share between the winning party or candidate and the second-placed party or candidate. So an election where the winner achieved 99 per cent of the vote would be considered relatively uncompetitive, whereas an election where two parties each got approximately 50 per cent of the vote would be considered highly competitive. Our measure might be reliable, but we might not say it is valid, since it doesn't capture the full definition of democracy.

One of the main challenges in comparative research, particularly large-N comparative research, is to do with establishing **equivalence of meaning**. That is, whether the theoretical concepts and empirical indicators of those concepts mean the same things and measure the same things in the different contexts to which they are applied. For example, the concept of populism may mean different things in different contexts. Thus, when we talk about populism in Latin America, we may be talking about something rather different to when we talk about populism in Europe. Different understandings of a concept can lead to different measures to tap into that concept (see Adcock and Collier 2001), which in turn can mean that even though we think we are talking about (and comparing) the same thing, we are actually talking about two very different things. Under these circumstances, the comparison is no longer particularly meaningful, because we are no longer comparing equivalent phenomena.

When we are only comparing a few countries, it is possible to explore these issues in some detail, and to take steps to ensure that there is a close match between concepts and measures in each of the cases being analysed. However, when we compare many countries,

we face a number of problems. On the one hand, our concepts may mean different things in different contexts, and so our measures are not valid. On the other hand, in order to establish equivalence our concepts and measures may become so broad, if not stretched (Sartori 1970), that their meaning becomes diluted. On the dangers of **concept-stretching**, see Chapter 5. It is only with careful analysis and detailed substantive knowledge that we can have confidence that our measures mean the same things in different countries.

Conclusions

In this chapter we have examined a number of different ways of doing comparative research. Broadly speaking, we can distinguish between these approaches according to the number of cases that are compared, and how the cases are selected. Comparative research is therefore primarily defined as a method of case selection. Different approaches to case selection are associated with various different problems, and the approach that is preferred may therefore depend upon the purpose for which it is used.

In particular, there are two main problems. The first is to do with selection bias. The second is to do with equivalence of meaning. The main problem with small-N studies and single-country case studies is that there is the real risk of selection bias. The cases we select can influence the answers we get. This means that our findings might not be very reliable, and could just be an artefact of the cases we have selected. If we selected different cases, we would come up with very different conclusions. One way to get round this is simply to examine more cases. But this comes with its own problems. When there are many cases, it becomes more difficult to ensure that the variables under investigation are both conceptually and empirically comparable. To ensure comparability we can end up using very general concepts and measures, which become abstract and devoid of meaning, undermining the validity of the measures.

Although these problems affect all comparative research, they should not be overstated. With careful design, we can take steps to tackle them head on, and by combining approaches, we can attempt to integrate the strengths of one approach with the strengths of another. However, it is worth bearing in mind that these issues of case selection are only the first part of the research process, and having chosen a case (or cases) to study, you then need to think about how you are going to study it/them. These methods of data collection and data analysis are discussed in detail in the following chapters.

Questions

1. What are the problems with selecting on the dependent variable? Have these problems been overstated?
2. If the cases you select affect the answers you get, does this mean that small-N comparison is inherently less reliable than large-N comparison?
3. Under what conditions does it make more sense to adopt an MDSD than an MSSD?
4. If a topic is worth investigating, it is worth investigating comparatively? Discuss.
5. What is concept-stretching? To what extent can it hinder comparative research?
6. To what extent is a small-N comparative study a qualitative research design?

7. What is the difference between methods for case selection, data collection, and data analysis? Does what we do at one stage influence what we do at the next?
8. What are the strengths and weaknesses of small-N comparison and large-N comparison?
9. How does QCA differ from small-N research?

Guide to Further Reading

Brady, H. E. and D. Collier (eds) (2004), *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Lanham, MD: Rowman Littlefield).
Contains a selection of essays which describe the ways in which qualitative approaches can contribute to the study of political phenomena.

Geddes, B. (2003), *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics* (Ann Arbor, MI: University of Michigan Press).
Provides a detailed discussion on case selection and how the cases you choose for analysis can affect the answers you get.

George, A. L. and A. Bennett (2005), *Case Studies and Theory Development in the Social Sciences* (Cambridge, MA: MIT Press).
An important discussion of the place of case studies in social science methodology. Emphasizes the importance of within-case analysis, provides a detailed discussion of process tracing, and of the concept of typological theories.

King, G., R. Keohane, and S. Verba (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ: Princeton University Press).
One of the classic methodology books. Provides a very influential perspective on small-N research, and how it can be carried out systematically and scientifically.

Landman, T. (2008), *Issues and Methods in Comparative Politics* (London: Routledge).
Provides an engaging treatment of issues in comparative research, with many examples drawn from the fields of democracy, human rights, and conflict. The book examines how case studies, small-N comparisons, and large-N comparisons have been used to answer a variety of different questions in comparative politics.

Ragin, C. C. (1987), *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies* (Berkeley, CA: University of California Press).
This book introduces the principles of QCA and Boolean Algebra using crisp sets. It provides an informative and engaging introduction to QCA and how it can be used to answer a variety of questions in comparative political research.

Ragin, C. C. (2008), *Redesigning Social Inquiry: Fuzzy Sets and Beyond* (Chicago, IL: University of Chicago Press).
This book develops Ragin's earlier work on QCA, and provides an in-depth treatment of how to calibrate fuzzy sets. It contains lots of helpful information on how to construct truth tables and how to analyse them.

References

Achen, C. and D. Snidal (1989), 'Rational Deterrence Theory and Comparative Case Studies', *World Politics* 41(2): 143–69.

Adcock, R. and D. Collier (2001), 'Measurement Validity: A Shared Standard for Qualitative and Quantitative Research', *American Political Science Review* 95(3): 529–46.

Almond, G. (1966), 'Political Theory and Political Science', *American Political Science Review* 60: 877–8.

Almond, G. and S. Verba (1963), *The Civic Culture: Political Attitudes and Democracy in Five Nations* (Princeton, NJ: Princeton University Press).
Bennet, A. and C. Elman (2006), 'Qualitative Research: Recent Developments in Case Study Methods', *Annual Review of Political Science* 9: 455–76.
Benoit, K. and M. Laver (2006), *Party Policy in Modern Democracies* (London: Routledge).
Berg-Schlosser, D. (2007), 'Determinants of Democratic Successes and Failures in Africa', *European Journal of Political Research* 47(3): 269–306.
Brady, H. E. and D. Collier (eds) (2004), *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Lanham, MD: Rowman Littlefield).
Brass, P. (2003), *The Production of Hindu-Muslim Violence in Contemporary India* (Seattle, WA: University of Washington Press).
Braumoeller, B. and G. Goertz (2000), 'The Methodology of Necessary Conditions', *American Journal of Political Science* 44 (October): 844–58.
Budge, I., H.-D. Klingeman, A. Volkens, J. Bara, and E. Tanenbaum (2001), *Mapping Policy Preferences: Estimates for Parties, Electors and Governments 1945–1998* (Oxford: Oxford University Press).
Butler, D. and D. Stokes (1969), *Political Change in Britain* (London: Macmillan).
Campbell, A., P. E. Converse, W. Miller, and D. Stokes (1960), *The American Voter* (New York: Wiley).
Caramani, D. (2009), *Introduction to the Comparative Method with Boolean Algebra*, Quantitative Applications in the Social Sciences Series (Beverly Hills, CA: Sage Publications).
Collier, D. and J. Mahoney (1996), 'Insights and Pitfalls: Selection Bias in Qualitative Research', *World Politics* 49(1): 56–91.
Collier, D., J. Mahoney, and J. Seawright (2004), 'Claiming Too Much: Warnings about Selection Bias', in H. E. Brady and D. Collier (eds), *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Lanham, MD: Rowman Littlefield), 85–102.
Coppedge, M. (1999), 'Thickening Thin Concepts and Theories: Combining Large N and Small N in Comparative Politics', *Comparative Politics* 31(4): 465–76.
Coppedge, M. (2012), *Democratization and Research Methods* (New York: Cambridge University Press).
Cunningham, D. (2016), 'Preventing Civil War: How the Potential for International Intervention Can Deter Conflict Onset', *World Politics* 68(2): 307–40.
Dion, F. D. (1998), 'Evidence and Inference in the Comparative Case Study', *Comparative Politics* 30(2): 127–46.
Eckstein, H. (1975), 'Case Studies and Theory in Political Science', in F. I. Greenstein and N. W. Polsby (eds), *Handbook of Political Science: Political Science: Scope and Theory*, vol. 7 (Reading, MA: Addison-Wesley), 94–137.
Geddes, B. (2003), *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics* (Ann Arbor, MI: University of Michigan Press).
George, A. L. and A. Bennett (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.
Gerring, J. (2007a), *Case Study Research: Principles and Practices* (Cambridge: Cambridge University Press).
Gerring, J. (2007b), 'Is there a (Viable) Crucial-Case Method?', *Comparative Political Studies* 40: 231–53.
Gleditsch, K. S. (2007), 'Transnational Dimensions of Civil War', *Journal of Peace Research* 44(3): 293–309.
Goertz, G. and H. Starr (eds) (2003), *Necessary Conditions: Theory, Methodology and Applications* (New York: Rowman and Littlefield).
Hackett, U. (2015), 'But Not Both: The Exclusive Disjunction in Qualitative Comparative Analysis (QCA)', *Quality & Quantity* 49(1): 75–92.
Hackett, U. (2016), 'The Goldilocks Principle: Applying the Exclusive Disjunction to Fuzzy Sets', *International Journal of Social Research Methodology* 19(5): 551–74.
Hawkins, K. (2009), 'Is Chavez a Populist? Measuring Populist Discourse in Comparative Perspective', *Comparative Political Studies* 42: 1040–67.
Heath, O. and L. Tillin (2018), 'Institutional Performance and Vote Buying in India', *Studies in Comparative International Development* 53(1): 90–110.
King, G., R. Keohane, and S. Verba (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ: Princeton University Press).
Kitschelt, H., Z. Mansfeldova, R. Markowski, and G. Toka (1999), *Post-Communist Party Systems: Competition, Representation, and Inter-Party Cooperation* (Cambridge: Cambridge University Press).
Levy, J. S. (2002), 'Qualitative Methods in International Relations', in F. P. Harvey and M. Brecher (eds), *Evaluating Methodology in*

International Studies (Ann Arbor, MI: University of Michigan Press), 432–54.
Lewis-Beck, M. (1986), 'Comparative Economic Voting: Britain, France, Germany, Italy', *American Journal of Political Science* 30: 315–46.
Lijphart, A. (1971), 'Comparative Politics and the Comparative Method', *American Political Science Review* 65(3): 682–93.
Lijphart, A. (1977), *Democracy in Plural Societies: A Comparative Exploration* (New Haven, CT: Yale University Press).
Lijphart, A. (1999), *Patterns of Democracy* (New Haven, CT: Yale University Press).
Lipset, S. M. (1959), 'Some Social Requisites of Democracy: Economic Development and Political Legitimacy', *American Political Science Review* 53: 69–105.
Lipsmeyer, C. S. (2002), 'Parties and Policy: Evaluating Economic and Partisan Influences on Welfare Policy Spending during the European Post-Communist Transition', *British Journal of Political Science* 32(4): 641–61.
Mainwaring, S. and T. Scully (eds) (1995), *Building Democratic Institutions: Party Systems in Latin America* (Stanford, CA: Stanford University Press).
Mair, P. (1996), 'Comparative Politics: An Overview', in R. E. Goodin and H.-D. Klingemann (eds), *A New Handbook of Political Science* (Oxford: Oxford University Press), 309–35.
Michelutti, L. (2008), *The Vernacularisation of Democracy* (London: Routledge).
Moore, B. (1966), *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World* (Boston, MA: Beacon).
Naroll, R. (1966), 'Scientific Comparative Politics and International Relations', in R. Barry Farrell (ed.), *Approaches to Comparative and International Politics* (Evanston, IL: Northwestern University Press).
Paine, J. (2016), 'Still Searching for the Value-Added: Persistent Concerns About Set-Theoretic Comparative Methods', *Comparative Political Studies* 49: 793–800.
Parry, G., G. Moyser, and N. Day (1992), *Political Participation and Democracy in Britain* (Cambridge: Cambridge University Press).
Posner, D. (2004), 'The Political Salience of Cultural Difference: Why Chewas and Tumbukas are Allies in Zambia and Adversaries in Malawi', *American Political Science Review* 98(4) (November): 529–45.
Przeworski, A. and H. Teune (1970), *The Logic of Comparative Social Inquiry* (New York: Wiley).
Przeworski, A., M. E. Alvarez, J. A. Cheibub, and F. Limongi (2000), *Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990* (Cambridge: Cambridge University Press).
Putnam, R. (1993), *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton, NJ: Princeton 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: University of Chicago Press).
Ragin, C. C. (2008), *Redesigning Social Inquiry: Fuzzy Sets and Beyond* (Chicago: University of Chicago Press).
Rihoux, B. (2006), 'Governmental Participation and the Organizational Adaptation of Green Parties: On Access, Slack, Overload and Distress', *European Journal of Political Research* 45: S69–98.
Rihoux, B. and C. Ragin (eds) (2009), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques* (Thousand Oaks, CA: Sage).
Rogowski, R. (2004), 'The Role of Theory and Anomaly in Social-Scientific Inference', in H. Brady and D. Collier (eds), *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Lanham, MD: Rowman & Littlefield Publishers, Inc.).
Rokkan, S. (1966), 'Comparative Cross-National Research: The Context of Current Efforts', in R. L. Merritt and S. Rokkan (eds), *Comparing Nations: The Use of Quantitative Data in Cross-National Research* (New Haven, CT: Yale University Press).
Rose, R. (1991), 'Comparing Forms of Comparative Analysis', *Political Studies* 39: 446–62.
Sartori, G. (1970), 'Concept Misrepresentation in Comparative Politics', *American Political Science Review*, 64: 1033–63.
Skocpol, T. (1979), *States and Social Revolutions: A Comparative Analysis of France, Russia, and China* (Cambridge: Cambridge University Press).
Thiem, A. and M. Baumgartner (2016), 'Back to Square One: A Reply to Munck, Paine, and Schneider', *Comparative Political Studies* 49: 801–6.
van Biezen, I. and D. Caramani (2006), '(Non) Comparative Politics in Britain', *Politics* 26(1): 29–37.
Verba, S., K. Schlozman, and H. Brady (1995), *Voice and Equality: Civic Voluntarism in American Politics* (Cambridge, MA: Harvard University Press).
Veugelers, J. and A. Magnan (2005), 'Conditions of Far-Right Strength in Contemporary Western Europe: An Application of Kitschelt's Theory', *European Journal of Political Research* 44: 837–60.
Weyland, K. (2002), *The Politics of Market Reform in Fragile Democracies: Argentina, Brazil, Peru, and*

Venezuela (Princeton, NJ: Princeton University Press).
Wilkinson, S. (2004), *Votes and Violence: Electoral Competition and Ethnic Riots in India* (Cambridge: Cambridge University Press).
Yadav, Y. (2000), 'Understanding the Second Democratic Upsurge: Trends of Bahujan Participation in Electoral Politics in the 1990s', in F. R. Frankel, Z. Hasan, R. Bhargava, and B. Arora (eds), *Transforming India: Social and Political Dynamics of Democracy* (Delhi: Oxford University Press).
Yin, R. K. (2009), *Case Study Research: Design and Methods*, 4th edition (Thousand Oaks, CA: Sage).

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

10 Historical Research

Chapter Summary

This chapter discusses the distinctions that have defined a separation between historical research and social scientific research, and how these are being challenged by scholars seeking to envision a genuinely 'historical social science'. It then discusses a number of historical approaches that have gained salience in Politics and International Relations. Finally, the chapter addresses how historical data is gathered, assessed, and used. Among the questions the chapter addresses are the following:

- What are the key similarities and differences between the nature and purpose of historical and social science research?
- What are the basics of historical research?
- What forms of historical analysis are used in Politics and International Relations?
- How do we formulate and support plausible historical interpretations?
- How do we locate essential historical information and assess various types of sources?
- What special considerations need to be made in using documents, archival sources, and historical writing as data?

Introduction

'The "past" is never really "past" but continuously constitutive of the "present"' (Bryant 1994: 11); consequently, 'in order to explain the structures of contemporary societies, one must investigate their historical origins and development' (Deflem and Dove 2013: 561). Few scholars would disagree with these statements. However, beginning with the construction of social science disciplines in the nineteenth century, the basics of historical research and social scientific research developed to a significant extent in opposition to each other. The system of knowledge acquisition established by the construction of social science disciplines fragmented the study of the social world in a number of profoundly significant ways. Social, political, and economic phenomena were defined as distinct from each other and assigned to separate fields of study (to Sociology, Politics, and Economics, respectively). Study of the 'modern/civilized world' and of the 'non-modern world' was also separated by assigning the former to Sociology, Politics, and Economics, and the latter to Anthropology. Finally, the study of the past and of the present was separated: history

became the study of the past, while social science disciplines focused on developing a science of the present. Since then, the introduction of a variety of analytic and methodological conventions has worked to further reinforce the separation between historical and social science research.

However, in recent decades greater concern with the complexities of historical change and cultural diversity has generated a growing appreciation of the mutual interdependence of history and the social sciences. Of course, political researchers have always drawn on historical studies to find out what history can tell us about contemporary events, to develop typologies of political phenomena, or to account for different patterns of political and socio-economic development. But increasingly, researchers in our field are recognizing that **temporality** (context and timing) matters very much to our ability to explain political phenomena, and that there are tools that historical approaches offer that can help us to 'place politics in time', as Paul Pierson puts it in his book, *Politics in Time* (2004). By enabling us to incorporate a stronger temporal dimension into our analyses, tools of historical research can help us to develop a better understanding of many political phenomena including, for example, institutional change and durability, state formation, revolution, trajectories of economic development, and globalization. This recognition is part of a broader trend that many see as leading to a convergence of history and the social sciences with respect both to methodological orientation and substantive research themes. In Politics and International Relations, this movement towards convergence has evolved to the point that researchers in our field today are often not even conscious of the fact that they are using approaches that come from history.

The trend towards the convergence of historical and social science research is an exciting development. Tools that historians and social scientists have helped to refine in recent decades are strengthening our ability to explore how past structures and relations constrain and enable the present, and what in emerging trends represents elements of continuity and change. They are also opening up new avenues of inquiry for researchers in our field. In the next section, we discuss how historical approaches and methods are contributing to our understanding of political outcomes.

Historical approaches in Politics and International Relations

One of the prime drivers of the convergence of historical and social scientific research is the increasing concern among social scientists with 'the *temporal* structurations of social actions and processes' (Hall 2007: 82). Researchers in our field are increasingly drawn to historical approaches because they enable us to incorporate a stronger temporal dimension into our analyses. Analysis becomes temporal by situating events in their historical and cultural contexts, taking into account related series of events that unfold over time, and thinking explicitly about timing and sequence.

There are two forms of temporality. Temporality can be conceived of in terms of (1) a broad historical period (time as context), and also of (2) the sequential active unfolding of social action and events. Each of these forms underpins a set of methods. We begin with the first form of temporality, time as context, which is represented by various methods of comparison.

Comparative and historical research

There are various types of research that are comparative and historical. These types are usually distinguished from each other on the basis of the number of cases and length of time studied, as shown in Box 10.1.

These types of research will be broadly familiar to those of you who have read Chapter 9 on comparative methods. It will be useful, however, to review all of them, and also to offer a more extended discussion of comparative-historical analysis.

Historical events research involves the investigation of events that either have an important impact on subsequent developments or provide an opportunity for testing the implications of a general theory. A number of political researchers have used this type of research to explore how shared representations of past events construct and maintain social identity and contribute to the intergroup dynamics that generate conflict. For instance, James Liu and Denis Hilton (2005) explore aspects of historical representation to explain the different responses by America's closest allies to the attacks in New York and Washington on 11 September 2001. In another study, Liu, Sibley, and Huang (2014) examine the relative contribution of historical representations in shaping political attitudes to current political issues in New Zealand (towards biculturalism) and in Taiwan (towards independence and self-determination).

This sort of research is extended, in **historical process research**, by examining a series of events that happened over a longer period of time in order to understand how processes change over time. For example, this type of research is used in a study of Chile's radical shifts in policy towards foreign direct investment (FDI) (Salcedo and Akoorie 2013). The study argues that shifting ideologies, expressed as politicized events, result in attitudinal shifts towards FDI by multinational enterprises (MNEs). To examine the argument, it focuses on three periods of investment, and, for each, identifies (1) key political events, (2) dominant political ideologies, (3) characteristics of FDI, and (4) responses of MNEs. Another example is a study of how legitimate authority for environmental governance is achieved (Knight and Knox-Hayes 2015). The authors argue that legitimacy for environmental governance evolves through a process in which the discursive practices of the scientific community, the community of expert economists, and the public at large creates, respectively, scientific, economic, and emotive authority for environmental governance. To investigate this argument they conduct historical process research to order events, practices, language, and activities over time, and then use this data as a basis for a longitudinal analysis of the Australian carbon market from 2006 and 2012.

BOX 10.1 Types of Comparative and Historical Research

	Cross-sectional	Longitudinal
Single case	Historical events research (one case and one time period)	Historical process research (one case/many time periods)
Multiple cases	Cross-sectional comparative research (many cases/one time period)	Comparative historical research (many cases and many time periods)

A third type of comparative and historical research compares two or more social settings or groups (usually countries) at *one* particular *point in time*. This is **cross-sectional comparative research**. An example of this type of research is a study of the relationship between ethnic diversity and social cohesion. A number of empirical studies have documented a negative relationship between ethnic diversity and 'generalized trust' (i.e. trust in other members of society—as opposed to trust in family and close friends). But these studies are often based on one country. Thus, to further explore this relationship, Marc Hooghe and his colleagues undertook a cross-sectional comparative study of diversity and trust across twenty European countries (Hooghe et al. 2009). The authors found that, while for *individuals* ethnic diversity was negatively related to 'generalized trust', this was not true for the country as a whole.

Let us now turn to a type of research which combines historical process research and cross-sectional comparative research, and which has become a prominent research tradition in our field: *comparative historical analysis* (CHA). CHA is concerned with (1) the 'identification of causal configurations that produce major outcomes of interest within historically delimited contexts'; (2) historical sequences and 'the unfolding of processes over time'; and (3) 'systematic and contextualised comparison' typically involving a few cases (Mahoney and Rueschemeyer 2003: 13). It has been used to study a wide range of issues, as Box 10.2 shows.

CHA begins with a set of theoretical propositions and then uses cases to illustrate, test, and refine them. The selection of cases is guided by three types of comparative-historical analysis: **parallel demonstration of theory**, **contrast of contexts**, and **macro-causal analysis**. These represent three different logics of inquiry; but Theda Skocpol and Margaret Somers (1980) treat these as together forming a 'research cycle'. Figure 10.1 shows what each type of analysis does, and how each informs the others.

The first two of these types of CHA will be familiar to those who have read Chapter 9.

BOX 10.2 Comparative Historical Analysis in Politics and International Relations

The causes and consequences of revolutions	Goodwin 2001; Parsa 2000; Foran 2003; Tilly 2003; Richards 2009: De Fronzo 2015; Walzer 2015
The emergence of democratic and authoritarian regimes	Herbst 2000, Wood 2000; Acemoglu and Robinson 2005; Gandhi 2008; Mahoney 2010
Processes of state formation and state restructuring	Hui 2005; Hibou 2005; Doornbos 2006; Centeno and Ferraro 2013; Waldner and Smith 2015
Racial and ethnic relations and national identities	Hechter 2000; Daniel 2006; Cornell and Hartmann 2007, Oberschall 2007, Hooghe et al. 2009
Gender and women's rights	Charrad 2001; Htun 2003; Inglehart and Norris 2003; O'Connor 2015
The development of welfare states	Huber and Stephens 2001; Swank 2002; Gough and Wood 2004; Haggard and Kaufman 2008
Economic development and industrial policy	Amsden 2001; Pomeranz 2001; Kohli 2004; Rodrik 2007; Mahoney 2010, McMichael 2012

Types of CHA:	What each does:	How these are linked in a research cycle:
Parallel demonstration of theory	**Shows the applicability of a general theory across a variety of cases.**	Parallel comparative histories which can be used to develop theories. ↓
Contrast of contexts	**Sets limits to the scope or claims of an overly-general theory** by highlighting unique aspects which distinguish cases that appear to be *similar* to each other.	Contrast oriented studies which can then be used to shed light on the limits of overly general theories. ↓
Macro-causal analysis	**Makes causal inferences about macro-historical processes and structures**; generates new historical generalizations and thus theory.	Macro-analytic studies can then be constructed to test the validity of existing hypotheses and develop new ones. Since these require theories as a starting point for their inquiry, this points to the continued need for

Figure 10.1 Different types and logics of comparative historical analysis

Parallel demonstration of theory is comparable to the 'Most Similar Systems Design': it selects cases that are *different* from each other in order to show that a theory applies in many different cases. The aim is to show how a particular theoretical proposition plays out in different national and historical contexts. Consider, for example, a study that examines risk-taking in a number of historical cases. In this study, Rose McDermott examines a proposition drawn from prospect theory: that individuals tend to be risk-averse when they are in a good position (in the 'domain of gains'), and more likely to take risks when they are in a bad situation (in the 'domain of losses'). McDermott compares the predictions of the theory and actual decisions reached in four historical cases of US foreign policy decision-making involving decisions taken by Jimmy Carter and Dwight Eisenhower in the domains both of losses and of gains.

The second type of CHA is based on a *contrast of contexts* logic (comparable to the 'Most Different Systems Design'). Cases are selected in this type of comparison to show how a similar feature—an institution or process general enough to be applicable across different contexts—produces different results in decidedly dissimilar contexts. Consider, for example, a study of constitutional durability (Elkins, Ginsburg, and Melton 2009). The study argues that constitutional durability is preferable to frequent constitutional revision because it is associated with democracy, and that the durability of a constitution is associated with three characteristics—*flexibility*, *inclusion*, and *specificity*. The study first establishes, through statistical analysis, the existence of positive relationships between the three characteristics and

constitutional longevity, and between countries with more durable constitutions and democracy. It then uses a contrast of cases design to examine two different sets of countries: one that includes three cases that have experienced short-lived constitutions (the Dominican Republic, Haiti, and Thailand); and a second set that includes two cases that have experienced long-lived constitutions (Mexico and Japan).

The third type of analysis, *macro-causal analysis*, compares historical cases for the purpose of making causal inferences about macro-level structures and processes, such as classes, states, and revolution. A macro-historical comparative study has the following features (Little 2000: 92):

- It defines its scope over a *long time period* and/or a *large geographical range*.
- It undertakes to account for large structural characteristics, processes, and conditions as historical *outcomes*.
- It singles out large structural characteristics in the social order as central *causes* leading to the observed historical outcomes.
- It aspires to some form of *comparative generality* across historical contexts, both in its diagnosis of causes and its attribution of patterns of stability and development.

A frequently cited example of macro-causal analysis is Theda Skocpol's *States and Social Revolutions* (1979). Skocpol's influential study argues that social revolution is brought about by external factors such as military defeat that change the relationship of state organizations to domestic political and social groups. When states weakened by external events experience 'revolutionary crises', 'revolts from below' accomplish 'changes in class relations that otherwise would not have occurred' (1979: 23). Skocpol demonstrates the argument by identifying a small class of relevant cases (China, Russia, France), specifying the social variables to be empirically evaluated (state structure, land tenure systems, forms of military organization), and then determining whether there are credible causal sequences among these variables in the several cases.

The Research in Action Box 10.1 provides an example of how CHA can be used to develop an argument about the termination of war.

All of these comparative and historical methods are concerned, in particular (though not exclusively), with one of the two forms of temporality we previously identified: *time as context*. Contextualizing events enable researchers to identify the causal influence of certain events on others because it helps us to better understand the relations between those events. Analyses that focus on the second form of temporality (the sequential active unfolding of social action and events) also help us to identify causal configurations, by examining relations that are stretched out over time. This type of analysis and form of temporality is represented by a second set of methods and approaches, which we shall discuss, below.

The timing and sequence of political events

The second form of temporality we previously identified is temporality as the sequential active unfolding of social action and events in and through time and concrete places. This is represented by historical institutionalist analysis and its associated methodologies, and **event structure analysis** (ESA).

Research in Action 10.1 When and Why Do Wars End?

In *War and Punishment* (2000), Hein Geomans addresses one of the most important, and least understood, aspects of war: when and why do countries decide to end wars? His answer focuses on the role that domestic politics play in the termination of war.

Goemans (2000) argues that, throughout the course of a war, belligerents make assessments of their chances to succeed in their war aims based on what they learn from their engagements with the enemy. These assessments are a critical factor in determining the possibility of settling the hostilities. War settlement, he argues, is only possible when a 'bargaining space' develops—when both sides come to similar conclusions about the probable outcome of continued war as well as the advantages of ending the war now rather than later. But the decision by leaders to then terminate war, Goemans maintains, can vary greatly depending on the type of government or regime they represent.

He categorizes regimes according to one of three types: repressive and exclusionary (dictatorships), semi-repressive and moderately exclusionary (mixed regimes), and non-repressive and non-exclusionary (democracies). The leaders of each of these types of regime face different domestic costs for losing a war, and this leads them to respond differently to new information, and to end a war at different points. Leaders of *democracies* who lose a war may be turned out of office; but democratic norms will preclude their incurring any further penalty. Leaders of a *strong repressive regime* will control sufficient resources to be able to survive the loss of war by 'buying off' the military and other segments of society that keep them in power. But the leaders of *moderately repressive regimes*, lacking both democratic norms and the resources of autocracies, incur the greatest costs and thus fear loss in war the most. They will likely see little risk in continuing a losing war and so may increase their minimal war aims and fight longer in the hope of turning the tide. Wars in which these regimes are on the losing side consequently last longer.

Goemans (2000) uses a basic rational choice bargaining model to explore the conditions and reasoning behind these strategies and to generate hypotheses regarding the strategies leaders of different regimes will choose. He hypothesizes that a leader's decision to continue or stop fighting will be based, at least partly, on how the terms of settlement will affect their political future; that, as more information is being revealed about the costs to them of an on-going war, leaders will modify their goals; and that, depending on how the terms of settlement will affect their political future, they will either choose to continue the war or they will agree to a peace settlement.

He then tests these hypotheses with statistical analyses based on a cross-national data set that includes characteristics of countries and conflicts. The statistical analysis supports Geomans' argument by showing the general association between different kind of regimes and the tendency to terminate war; but it cannot firmly establish the causal direction which the argument states. Goemans therefore develops detailed historical case studies to explore how two non-repressive, non-exclusionary regimes (France and Great Britain) and two semi-repressive, moderately exclusionary regimes (Germany and Russia) changed their war aims in the First World War. This CHA, not only shows that different kinds of regimes will calculate the costs of continuing or ending a war differently; it also offers a new perspective on the First World War and why the war lasted for four years.

Historical institutionalism

Historical institutionalism is closely related to CHA. Both are concerned with the analysis of temporal processes, but historical institutionalism is concerned more specifically with bringing questions of timing and temporality to the centre of the analysis of *institutions*, something which is the concern of many, but not all, comparative historical analyses. And while there are many historical institutionalist studies that engage in systematic comparison, many do not, and so 'do not fall within the field of comparative historical analysis' (Mahoney and Rueschemeyer 2003: 11).

Historical institutionalism is not new. Many of the classics of Politics emphasize aspects of historical context to account for cross-national variation in institutional outcomes. A notable example is Alexander Gershenkron's *Economic Backwardness in Historical Perspective* (1962), which endeavours to show that the process of industrialization is essentially different for late developers than for early developers; and that, consequently, *when* a country industrializes necessarily affects *how* it industrializes. But historical institutionalism took a new turn with the 'new institutionalism' that was launched, in 1984, by political scientists James March and Johan Olsen. March and Olsen argue that institutions shape, and are not just shaped by, political behaviour and social change; and that they shape behaviour by giving meaning to interactions and providing the *context* within which interactions take place.

What is distinctive about the historical institutionalism that emerged from this new thinking is the conceptual tools it has developed for exploring questions of timing and temporality in 'the creation, reproduction, development, and structure of institutions over time' (Fioretos 2011: 372). Among the most important of these concepts is that of 'critical junctures' and the associated concepts of 'path dependence' and 'positive returns' or feedback. A **critical juncture** is an event which sets in motion courses of action that become difficult to reverse. As Paul Pierson points out, 'relatively small events', if they occur 'at the right moment', can have 'large and enduring consequences' (Pierson 2004: 44). This is because, once institutions begin to move in a particular direction, **positive feedback** generates self-reinforcing mechanisms that help to prevent further change by removing certain options from the subsequent menu of political possibilities and generating irreversibilities (Pierson and Skocpol 2002: 699). This self-reinforcing persistence of institutional structures is called **path dependence**. These concepts help to explain how and why institutions originate (critical junctures), persist, and evolve (path dependence).

The launch of the new institutionalism generated a number of other approaches, in addition to historical institutionalism. Three of these, in particular, have become popular: *sociological* (or normative) institutionalism, *discursive* institutionalism, and *rational choice* institutionalism. While all of these are used to analyse political outcomes, *historical* institutionalism differs from them in the substantive scope of the questions it addresses, and in its understanding of what shapes actors' preferences and actions (Fioretos 2011: 372–3). Both the sociological and the discursive traditions emphasize, though in somewhat different ways, the role of socialization, socially constituted norms, and meaning systems in shaping individual preferences and action. Rational choice approaches assume that actors' preferences and action are shaped by end-point comparisons: a calculation of the *prospective* alternative that is less costly or more efficient. In contrast to sociological and discursive traditions, historical institutionalism gives less weight to the role of socialization and norms in shaping preferences. And in contrast to rational choice approaches, it focuses on how institutions develop and adapt rather than on how they function; and it assumes that preferences are informed by comparing the costs and benefits of adapting to new circumstances with those of maintaining or losing their investments in *past* arrangements.

Researchers in Politics and International Relations have used historical institutionalism to explore elements of persistence and change in the modern state, political parties, and organized labour; to examine the origins of state sovereignty, global orders, and international organizations; and to study developments in international security cooperation, international law, trade, finance, and other areas of regulation.

Let us consider an example of a historical institutionalist analysis of institutional persistence and change which addresses an intriguing puzzle: why we find cases in which political institutions change but the underlying economic institutions persist. James Robinson argues that the persistence of *economic* institutions in the face of change in *political* institutions can be explained by how elites form around sets of institutions and how they persist and reproduce over time. He argues that 'once an elite has the power to determine the choice of institutions this power will tend to persist over time' (Robinson 2010: 9). Why this is can be explained by distinguishing between two types of power: *de jure* power, which includes 'all types of power that originates from the political institutions in society', and *de facto* political power—power that is not allocated by political institutions, but is achieved by coercion, cooptation, or other means (Robinson 2010: 6). Groups possessing some combination of these types of power will be able, not only to choose economic institutions that reproduce their *de facto* power, but also to choose political institutions in the future that will reproduce their *de jure* power.

Robinson applies the argument to two cases in which economic institutions persisted in the face of change in political institutions. The first is the case of the US South where, following its defeat in the Civil War, significant changes took place in political institutions (the abolition of slavery and the enfranchisement of freed male slaves) but where 'a labor-intensive, low wage, low education and repressive economy' emerged 'that looked remarkably like the antebellum Southern economy'. The reason, Robinson argues, was that while white elites lost much of their *de jure* political power, their *de facto* power (influence over the legal system and control over land and the local political system) enabled them to block economic reforms, extract rents, and repress the labour force (2010: 10). It was only with increased black outmigration and the rapid spread of mechanized cotton picking after the Second World War that the ability of the landed elite to exercise *de facto* power in these ways declined.

The second case focuses on a question which has drawn much attention from political analysts and observers: why South Africa's transition from white rule to democracy in 1994—the election of the African National Congress (ANC) and redistribution of political power towards the black majority—failed to bring about large economic changes. Robinson argues here again that while the change in political institutions significantly redistributed *de jure* political power, the white economic elite's *de facto* power enabled it to continue to influence the nature of economic institutions. Beginning in 1993, white firms put members of the ANC elite on boards of directors and sold them discounted shares to the consortia they had formed. This greatly enriched the politically powerful new black elite, and gave them a vested interest in the set of institutions—the highly concentrated industrial structure and entry barriers—that had previously generated huge rents for the formerly all-white elite. So, when political institutions changed, the ANC did not adopt policies in line with the redistributive policies that it had advocated for decades, but instead adopted market-friendly policies which have substantially increased income inequality.

A second example of a historical institutionalist analysis is one which addresses another puzzle, this one relating to US policy towards the Middle East peace process (MEPP). It is commonly assumed, as Roland Dannreuther points out, that US policy towards the peace process is driven by strategic competition, by the power of domestic lobbies (i.e. by the 'Israel lobby'), or by a mixture of both. But Dannreuther argues that, while these factors

are certainly part of the story, they fail to adequately explain the puzzle of why the US has given consistent and almost unconditional support to Israel despite the strategic problems this creates for broader US Middle East policy. Dannreuther proposes that we can gain insight into this puzzle by treating the MEPP as an institution and exploring how key historical junctures shaped the approach that the US adopted to it. The critical juncture, he argues, was Henry Kissinger's 'distinctive and innovative approach' to the MEPP during the 1973 war, which subsequently became entrenched and reproduced (2010: 189). Kissinger's approach consisted of three principal components: (1) the assumption that 'peace would come from the strength of the US–Israeli relationship, which 'would draw the Arabs into negotiating directly with the US' while leaving the Soviet Union on the sidelines (2010: 201–2); (2) step-by-step 'shuttle' diplomacy that placed the US at the centre of the web of bilateral negotiations between Israel and the Arab states; and, as an unintended consequence of the first two components, (3) granting substantial concessions to Israel in exchange for its cooperation. Because this approach subsequently became '"locked in" by a number of positive feedback mechanisms', subsequent US mediators found that the approach to and conceptualization of the MEPP that Kissinger had forged at this critical juncture in the Arab–Israeli conflict 'proved institutionally "sticky" and difficult to diverge from' (2010: 202–3).

These are examples of how researchers in Politics and International Relations use concepts drawn from historical institutionalism (critical junctures, positive feedback, and path dependence) to explain institutional continuity or persistence. In the first example, James Robinson highlights the potential role of elites in institutional persistence through the use of case studies, and the construction of a **narrative** account (for each case) that links a cause (elite power) to institutional persistence. In the second example, Roland Dannreuther constructs a narrative account that offers what he calls an 'alternative interpretive lens' for explaining US policy towards the Arab–Israeli conflict, one that is 'more historically sensitive and less analytically deductive' than other explanations (2010: 194). Both studies are suggestive of the role of factors that are, perhaps, underemphasized in other accounts. However, neither study links the hypothesized causes (i.e. a critical juncture or elite power) to institutional continuity. Nor do these studies explain how the evidence they employ bears on their argument in relation to competing explanations.

This is probably a good place to point out that, while historical institutionalism provides tools that help researchers explore 'the temporal structurations of social actions and processes' (Hall 2007: 82), it does not, in and of itself, provide a method for establishing a *causal* link between hypothesized causes and outcomes as, for instance, CHA does (through the use of systematic and contextualized comparison of a small number of cases). Historical institutionalists tend to use CHA as a method. They also tend to favour a method we have not yet discussed: process tracing.

Process tracing

Process tracing is a method for identifying the causal relations that connect hypothesized causes and outcomes. It entails carefully tracing the sequence of events constituting a process (e.g. institutional persistence or change) in order to identify the underlying causal mechanisms that link a variable (e.g. elite persistence or a critical juncture) with an

outcome. It is therefore a good method for exploring the events and mechanisms that constitute path dependent historical processes.

An analysis of path dependence requires identifying both *events* and *mechanisms* (intentions, determinants). The event will be a critical juncture: a point in time 'where, for some reason, an old path is abandoned and a new path is entered upon' (Bengtsson and Ruonavaara 2010: 8). Identifying a critical juncture requires that the analyst first focuses on a point in time where some otherwise plausible paths are not considered or taken (T_2), and then goes backwards in history to the situation (which might not have been particularly noted at the time) in which the closing off of alternative paths occurred (T_1). In the study of the MEPP that we discussed in the section on historical institutionalism, Dannreuther (2010) does this by first addressing a 'puzzle' relating to the current US policy of consistent and almost unconditional support for Israel, and then going backwards in history to the 1973 war which, he argues, constituted a critical juncture in the institutionalization of US policy towards the peace process.

However, establishing a causal link between this event and the outcome in question requires additional steps. The analyst must (1) identify some sequence of relevant occurrences that took place in the interval between the event (T_1) and the outcome (T_2); (2) break the chain of occurrences or decision-making process into smaller steps; and then (3) look 'for observable evidence of each step' (Van Evera 1997: 64). The analyst can look for this evidence in histories, archival documents, interview transcripts, or other sources relating to the case; they can also examine these sources for evidence of underlying causal *mechanisms* or processes—evidence that reveals, for instance, the motives or perceptions of government officials or decision-makers. Finally, the analyst should establish how different pieces of evidence bear on competing explanations.

It should be remembered that to confirm a hypothesis or argument requires that the analyst address, or control for, alternative explanations. This is a key component in process tracing. Andrew Bennett and Jeffrey Checkel suggest that researchers employing process tracing should first *construct a list of potential alternative explanations*, including theoretical explanations in the academic literature, more context-specific arguments, and 'the implicit theories of journalists or others following the case, and the understandings participants have about what they are doing and why they are doing it' (Bennett and Checkel 2014: 18). They should then consider the probative value (degree of proof or value) which that evidence has for alternative explanations.

We have said that process tracing traces the sequence of events constituting a process to establish the causal relations among them and identify the underlying causal mechanisms that produce the outcome they are assumed to produce. There is a second, more systematic and formalized, method for converting a historical narrative into an analytical causal explanation—ESA—to which we now turn.

Event structure analysis

There is a tendency in narrative construction to assume a kind of 'genetic determinism': to assume that earlier happenings determine what comes later (Skocpol and Somers 1980: 193). Of course, the fact that one event happened first doesn't necessarily mean that it caused all subsequent events; but narratives often fail to make explicit the causal logic of sequences.

ESA is an analytic procedure designed to probe the temporal sequences in the narratives that analysts construct (Griffin and Korstad 1998: 145), by helping the analyst to 'unpack' an event—that is, breaking it into its constituent parts—'and analytically reconstituting it as a causal interpretation of what happened and why it happened as it did' (Griffin 2007: 5). There are three steps involved in this, which we discuss later in this section. Though the steps involved in the analysis can certainly be conducted without it, there is a computer program, ETHNO23, that aids researchers employing ESA.[1] What this program does is to ask the researcher, first, to construct a list of events that comprise the narrative and enter them into the program as raw data; and, second, to answer a series of yes/no questions concerning the relations between the series of events.

Larry Griffin and Robert Korstad illustrate how this works using a case of a successful attempt by the United Cannery, Agricultural, Packing, and Allied Workers of America union (UCAPAWA) to organize African-American industrial workers at the R. J. Reynolds (RJR) tobacco plant in Winston-Salem, North Carolina between 1941 and 1944.

In 1941, 'RJR operated the largest tobacco manufacturing facility in the world, employing 12,000 unorganized', mostly African-American, workers of both sexes. 'RJR had long manipulated racial and gender divisions among its workers to thwart unionization' (Griffin and Korstad 1998: 148), so this organizing drive represented an extraordinary victory, and, in particular, for that time and place. Moreover, given the persistent challenges faced by unions and by African-Americans in the US, the case—as the authors point out—merits close analysis. Here are the steps involved in ESA in relation to this case.

Step 1: Construct, in narrative form, a brief account of what happened.

The UCAPAWA moved into Winston-Salem to organize RJR's black workforce in late 1941. Using its connections to African-American organizations, UCAPAWA organizers met with African-American workers and, by 1943, were able to set up an all-black Tobacco Workers Organizing Committee (TWOC). The rising cost of living and a speed-up of production in the summer of 1943 intensified the chronic resentment amongst RJR's black workers. On 17 June, and as a result of the speed-up, an African-American worker was belittled by her supervisor on the shop floor for being behind in her work. In response, 'black women workers, some of them TWOC members, began to discuss the possibility of a protest strike'. At the same time, 'another African-American worker died on the shop-floor, reportedly after his white foreman refused to let him go home or get medical attention'. TWOC women then spontaneously staged a sit down strike, which quickly spread. Though the UCAPAWA did not call for a large strike, other African-American workers then also walked off their jobs in solidarity, and, after sustained work stoppage, about 6,800 of them (perhaps two-thirds of RJR's black workforce) joined the TWOC (Griffin and Korstad 1998: 148).

Step 2: Convert this narrative into a series of short statements to be entered into ETHNO, each expressing an action and consisting of one line of text. After reducing the narrative in this way, the author then entered the following nine statements into ETHNO:

1. UCAPAWA targets RJR's African-American workers for unionization.
2. African-American workers meet with UCAPAWA organizers about unionization.
3. UCAPAWA forms all-black TWOC to include strategically placed workers.
4. RJR speeds up production process.

5. White foreman abuses African-American female for being behind in her work.
6. African-American dies on shop floor allegedly after white foreman refused to permit him either to leave or to get medical assistance.
7. African-American female workers sit down, stopping production.
8. Thousands of African-American workers strike.
9. Almost 7,000 African-American workers join TWOC.

When a series of statements are entered into the program, ETHNO abbreviates the action word in each of them to be used for the diagram it will eventually produce (e.g. meet = Mee, target = Tar). It then asks the researcher a series of yes/no questions on the relations between events. It starts by asking whether event #1 is required for the occurrence of event #2. The exact wording is: 'Does "Mee: African-American workers meet with UCAPAWA organizers about unionization" require "Tar: that UCAPAWA/CIO target RJR African-American workers for unionization" (or a similar event)?' It might be assumed that, having targeted these workers for unionization, the organizers 'naturally' would have set up a meeting with them. On reflection we can see that the existence of a targeting strategy is not *sufficient* to bring about a meeting, so the researchers' answer to the question might initially have been 'no'. However, the question forced them to think about a host of issues they would otherwise not have considered. For instance, they considered the question ETHNO asked in terms of a counterfactual: 'Would the organizers have met with African-American workers had they not targeted them for unionization?' This led them to conclude that, though the targeting was not sufficient for bringing about the meeting, it was *necessary* to bringing it about because, had the organizers not targeted these workers previously, there would not have been union organizers for African-American workers to meet with, nor a demand for them from the rank-and-file. Consequently, their final answer to this question was 'yes'. ETHNO then produced a diagram that indicated the existence of a direct causal tie between the two actions (see Figure 10.2).

ETHNO continued its interrogation to determine the linkages between the remaining statements. The researchers answered 'yes' to the next question, about the relationship between 'Mee' (the meeting) and 'For' (the formation of the TWOC), but then answered 'no' to the next five questions: the relationships between 'Spe' (the RJR speed-up) and the three prior events in the chronology ('Tar', 'Mee', and 'For'). They also answered 'no' to queries about whether the foreman's abuse of a slow worker ('Abu') was causally related to both 'For' and 'Mee'. They answered 'yes' to the next question: was 'Spe' necessary for 'Abu'? After the researchers answered these questions, ETHNO's diagram is shown in Figure 10.3.

Figure 10.2 ETHNO diagram of causal relations I

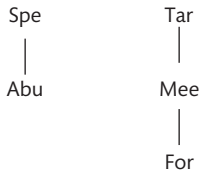


Figure 10.3 ETHNO diagram of causal relations II

Figure 10.3 depicts two parallel streams of action: it shows, on the one hand, events which resulted in putting a union organizing committee in place; on the other, it shows the 'speed-up induced abuse of an African-American worker by her white supervisor'. At this point, the two streams are not connected. But some future act might alter the significance of both streams and produce 'an effect impossible for either singly to induce' (Griffin and Korstad 1998: 153).

In fact, after the researchers had answered all the remaining questions, ETHNO's diagram of how the nine actions constituting the chronology were causally related indicated a third stream of causally connected events which linked the death of the worker ('Die'), the sit-down strike of female African-American workers ('Sit'), the strike involving thousands of African-American workers ('Stk'), and the culminating event in which nearly 7,000 African-American workers joined the TWOC ('Joi') (see Figure 10.4).

What this shows is that TWOC's growth ('Joi') was not the inevitable outcome of UCAPAWA's organizing strategy ('Tar') or the events which followed from it ('Mee' and 'For'). In fact, the TWOC might have remained ineffectual had the sit-down strike not occurred. But neither the death of the worker nor the speed-up of production and the abuse of a worker on the shop floor was *sufficient* to have predicted the sit-down strike. As the diagram shows, it was the confluence of (1) the formation of the TWOC, 'which provided

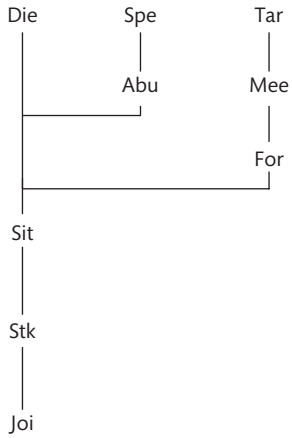


Figure 10.4 ETHNO diagram of causal relations III

BOX 10.3 Key Terms

Temporality	The historical context, and development and unfolding over time of events, processes, and relationships.
Historical events research	Investigation of an event that has an important impact on subsequent developments or provides an opportunity for testing the implications of a general theory.
Historical process research	Examination of a series of events in order to understand how processes change over time.
Cross-sectional comparative research	Research which compares two or more social settings or groups (usually countries) at *one* particular *point in time*.
Comparative historical analysis	Investigation that begins with a set of theoretical propositions and then uses multiple cases and time periods to illustrate, test, and refine them.
Parallel demonstration of theory	CHA (comparable to 'Most Similar Systems Design') that selects cases that are *different* from each other in order to show how a particular theoretical proposition plays out in different national and historical contexts.
Contrast of contexts	CHA (comparable to the 'Most Different Systems Design') that selects cases to show how a similar feature—an institution or process general enough to be applicable across different contexts—produces different results in dissimilar contexts.
Macro-causal analysis	CHA, typically focused on a long time period and/or a large geographical range, concerned with making causal inferences about macro-level structures and processes, such as classes, states, and revolution.
Historical institutionalism	Analysis concerned with bringing questions of timing and temporality to the centre of the analysis of *institutions*.
Process tracing	A method that entails carefully tracing the sequence of events constituting a process (e.g. institutional persistence or change) in order to identify the underlying mechanism that connects hypothesized causes and outcomes.
Event structure analysis	An analytic procedure that helps an analyst to break an event into its constituent parts in order to develop a causal interpretation of what happened and why it happened as it did.

both the organizational base for collective action and an interpretive frame molding and amplifying the racial, gender and class grievances of those who spontaneously stopped production'; (2) RJR's speed-up of the production pace, and two racially charged 'triggering' incidents; (3) the abuse; and (4) the death of an African-American worker. In showing us why the sit-down strike and the mass organization of workers happened, and why they happened when they did, ESA helps us 'to grasp how unionization as a structural possibility was transformed by social action into unionization as a lived reality' (Griffin and Korstad 1998: 153).

We have shown how the methods and approaches we have discussed in this section place temporality at the centre of the analysis of political phenomena and how, by doing this, they

contribute to our understanding of outcomes of interest to political researchers. We have discussed how historical and comparative methods help us to discover the causal influence of certain events on others by contextualizing events and, in this way, enabling us to understand the relations between those events. We have also discussed how historical institutionalism, by highlighting the unfolding of processes over time, and using methods such as process tracing, enables us to identify causal configurations and mechanisms. ESA takes this further by enabling researchers to probe the links between the actions that constitute the chronology of a given event for potential causal ties among those actions. These methods combine features drawn both from formal social science methodology and from the various ways that historians infer causality and meaning from events. For instance, we can see that CHA, in contrast to much historical research, goes beyond 'the descriptive narrative and the particularity of single events . . . to seek more general insights'; and, in contrast to much conventional social science, it is 'case-sensitive rather than variable-driven' (Anheier and Katz 2006: 289). It is also evident that ESA thoroughly integrates the theoretically/empirically general and the historically particular, and merges explanation and interpretation.

Something which all these methods and approaches have in common is their reliance on careful interpretation of reliable sources. The next section considers what these sources are, where they are, and how to evaluate them. First, however, take a look, in Box 10.3, at the key terms that this chapter has introduced so far.

Using documents, archival sources, and historical writing as data

Identifying sources of data in historical research

In identifying sources of data in historical research, it is usual to distinguish between *primary* and *secondary* sources, with documents and archival material of various sorts falling in the former category, and historical writing belonging to the latter.

Primary sources are materials written or collected by actors who actually witnessed the events they describe. They provide direct or first-hand evidence about an event, object, or person, and show minimal or no mediation between the document/artefact and its creator. Participants in the political processes we are concerned to understand—including civil servants, members of advisory councils, and representatives of pressure groups involved in decision-making processes—generate official records, such as party programmes, parliamentary proceedings, resolutions, speeches, treaties, press conferences and press reports, television and radio interviews, and correspondence. Other examples of primary sources include letters, manuscripts, diaries, journals, newspaper and magazine articles (factual accounts), speeches, interviews, memoirs, documents and records produced by government agencies, recorded or transcribed speeches, interviews with participants or witnesses of an event, or with people who lived during a particular time, photographs, maps, postcards, posters, audio or video recordings, research data, and objects or artefacts such as works of art, buildings, tools, and weapons.

A great deal of political research relies on **secondary sources**. These are materials produced some time after an event happened, and so contain information that has been

interpreted, commented on, analysed, or processed in some way. These sources include biographies, histories, encyclopaedias, newspaper articles that interpret, or journal articles and books written by social scientists. Historians have traditionally seen primary sources as closer to 'the truth', and so superior to these secondary sources. But all sources are partial (or 'biased') in some way. Consequently, all sources—primary or secondary—must be critically assessed to avoid bias, errors, omissions, and exaggerations. We will discuss how to evaluate sources further on in this section. First, however, we discuss where to locate these sources.

Where to locate sources

Primary sources

These are kept in museums, archives, libraries, or private collections; but an increasing number of primary sources, from all over the world, have been digitized and are available online.

Archives (or record offices) maintained by national and local governments contain a wide range of official records, as well as considerable quantities of 'private' material of value to researchers. In most countries, the largest body of these records is found in *national state or government archives*. But there are also *local government agencies* and county and other various local public record offices which contain an enormous amount of published and unpublished, public and primary source material, including collections of official documents, and records of local businesses, estates, and associations. *Libraries* also hold collections of published material of international, national, and local interest. Local studies libraries and university libraries contain official reports, and family histories, autobiographies, and memoirs, as well as oral histories—interviews and video memoirs—that provide important perspectives for researchers. Private associations can also be a very valuable source for documents, such as records of guilds, trade unions, political parties, pressure groups, families, and businesses. Among these, we might also include church records, which many historians have found to be a rich source of information on aspects of local life not recorded in other types of sources.

Given the enormous amount and variety of primary source material available to researchers, identifying relevant primary sources may seem a daunting task. However, there are plenty of tools to help us identify primary source material. One widely used tool is *WorldCat*, a database of more than a billion records of books, archival materials, maps, visual resources, and other materials held in more than 17,000 academic, public, special, and national libraries around the world. Of course, the websites of many repositories can be searched directly. *ArchiveGrid*, which helps researchers look for primary source materials held in archives, libraries, museums, and historical societies, contains over four million records that describe materials held by some 1,000 different institutions. *Archive Finder* combines the cumulative index to the *National Inventory of Documentary Sources in the UK and Ireland* (NIDS UK/Ireland) and *ArchivesUSA*, a directory of over 5,500 repositories and more than 161,000 collections of primary source material across the United States.

Secondary sources

Take a look at the discussion in Chapter 4 for suggestions about how to begin your search for secondary sources. In addition to those suggestions we want to point out that there are numerous reference sources that can help you to locate relevant secondary sources.

You might start with online encyclopaedias, both specific entries and the bibliographies or reference lists at the end of the entries. Books relevant to your research can be identified with the help of online library catalogues; and articles can be found with the many, many article databases that exist, such as JSTOR and *Historical Abstracts*. There are also very many newspaper databases, such as *Lexis-Nexis* and *News Source*, which can enable you to search international newspapers, newswires, and TV and radio news transcripts.

Evaluating sources

All sources, whether primary or secondary, must be assessed critically; but each of these types of sources entails specific considerations.

Primary sources

Primary sources must be evaluated in order to establish (1) their authenticity, and (2) the reliability or accuracy of the information they contain.

Establishing the authenticity of a primary source requires the researcher to engage in what historians call '**external criticism**': an assessment of the *external characteristics* of a source, including *when*, *where*, *why*, *by whom*, and *for whom* it was produced. This enables the researcher to evaluate the context and biases which inform the content of a source. Knowing *when* and *where* the source was produced helps us understand it in the context of what was going on at the time. Knowing something about *why* the source was produced, and under what circumstances, is also important. Was the source produced in order to record events, to deceive readers, or to persuade them to adopt a certain perspective or position? Knowing something about *who* created the source, and *for whom*, can also help us explore its biases or overall angle of vision.

Researchers must also determine the truthfulness, accuracy, and evidential worth of a primary source through what historians call '**internal criticism**'. This requires us to find out something about how the author of the source is related to its subject matter. Can the author be considered a trustworthy observer? How competent was the author to report critically and with comprehension?

We noted that an increasing amount of primary source material is being made available to researchers online. This is vastly increasing our access to primary source materials. However, there are some additional considerations involved in using these sources. Unlike print sources, which have editors to oversee the quality of their content, much information published on the Web has not been formally reviewed. Consequently, to use material on the Web you will need to evaluate the sources and the sites where it is posted. This evaluation should be based on the following four criteria (source: http://umf.maine.libguides.com/ld.php?content_id=10843324):

1. **Authority:** Is there information about the document on the website where you have accessed it that enables you to assess how authoritative the source is? For instance, does the website tell you what *the origin of the document or artefact* is and *who created it*? Does it tell you *where the original is located*? What do you know about the website that has posted the document? What type of website is it (.com, .edu, .gov, .org, etc.)? What do others say about the site? (You can check for reviews at *The Internet Scout*

Report (http://scout.wisc.edu/); and *The Public History Resource Center* (http://www.publichistory.org/reviews/rating_system.html).) Do other sites link to it? (Google has a feature that allows you to check this (go to http://www.google.com/help/features.html#link).)

2. **Accuracy:** Does the document that is posted accurately represent the contents of the original? Is it a scan, transcription, or an image of the original? Is the original available for comparison? Does the document make sense in light of other available information? Are there references and/or links to related collections or descriptive information?
3. **Accessibility:** How open or closed is access to the site? Is special software required to view the contents of the site? If so, is that software freely available or must it be purchased? Is there a fee for viewing documents or other items?
4. **Purpose:** Who is responsible for the website, and why have they posted this document on this site? Is there an 'about' link to information about the institution, and if so does that information indicate a bias or an agenda? Is the content of the site balanced? Are materials used in a way that is meant to present a particular point of view? Are documents omitted that you would expect to be included?

Secondary sources

There are also particular issues that arise for political researchers who rely on secondary sources. These must be evaluated to determine the extent to which they can be treated as reliable reports of the past. Do they contain errors, biases, exclusions, and exaggerations? Do they reflect the author's personal commitments or implicit theories of or perspectives on human behaviour? Here are a few suggestions to help you to think about these questions.

First, it is important to think of different historical accounts, not as offering different sets of *facts*, but as a series of *competing interpretive narratives* which are the focus of scholarly debate. To think about different accounts in this way, it is best to start with the most recent scholarship and read backwards (since historical controversies are not about basic information—e.g. 'Barack Obama became president of the US on 4 November 2008'—but about different interpretations of their underlying causes and consequences). Second, rather than relying on a single historian's account, we should consider the accounts of a number of historians who differ methodologically and substantively, and look for an overlap in them. Third, we need to be attentive to the perspectival and theoretical commitments that sources carry, which requires us to know something about the intellectual heritage and theoretical dispositions associated with each scholar.

So if different authors offer vastly different interpretations of the same historical events, how are we to choose from among the available accounts? There are two issues that must be considered.

The first issue centres on how to choose from the available sources in a way that avoids 'selection bias'—that is, a bias towards those accounts which fit with the argument being investigated and defended. The danger of this type of bias arises whenever we must choose among conflicting, or partially conflicting, available historical accounts. The 'nub of the issue', as Ian Lustick points out, is that the search for available studies 'may well entail . . . a heavy selection

bias toward works by historians using implicit theories about how events unfold and how people behave very similar to the theory under consideration by the social scientist' (Lustick 1996: 607). A second issue relates to **confirmation bias**: the tendency of the researcher to limit the search for evidence in secondary sources to only those that provide confirmation for the researcher's argument. This type of bias arises when a researcher approaches research as a process of gathering as much evidence as possible to support a particular hypothesis or point of view, and ignores any information that weakens it.

There are strategies to help you to avoid these biases. First, include an analysis of patterns across the range of historical accounts relating to your topic. Second, look for regularities that appear across accounts that are otherwise different or contradictory—that is, based on different approaches or archival sources, or which develop different perspectives or reach different conclusions. Third, note alternative versions or sources that contradict those on which you rely, that is, those that tell a different story. These 'strategies' are all directed towards helping you to *justify your choices*. You can reduce the amount of additional space this might take up by using these techniques to address only those 'elements of the background narrative that are either particularly controversial in existing historiography or that are particularly salient for the theoretical argument under review' (Lustick 1996: 614).

As in every choice made in the research process, our choice of sources always depends on what is required for developing a meaningful and persuasive investigation of our research question. But the selection of source material requires us to be critical and rigorous about how we evaluate the research of others. In using documents, archival sources, and historical writing as data, this requires us both to read historiography critically and to engage in some independent inspection and interpretation of the materials from which historians have drawn inferences. Again, the overall point is that researchers must be self-aware and critical about every choice they make in the research process, and make the considerations that enter into those choices clear and transparent to others.

Conclusion

Political research and history have always come together in key areas of inquiry in our field: in the study of the origin and development of states and political institutions, the requisites of democracy and of development, and the dynamics of social classes and of revolution, as well as in international security, international political economy, and foreign policy analysis. In recent years, however, growing recognition of the mutual and necessary interdependence of history and the social sciences has led to the increasing convergence between historical and social scientific methods. Researchers in our field are increasingly using tools from historical approaches and methodologies, including CHA, historical institutionalism, process tracing, and event sequence analysis, in order to develop a fuller and more systematic understanding of the context and *temporal* structuration of actions and events, and of multiple, time-ordered and conjunctural causality.

This is an exciting development in the field. Efforts to develop a sound understanding of the origins and nature of the contemporary political order, and of the nature and direction of current trends of change, depends on our ability to situate events in their historical and

cultural contexts, trace their precise unfolding, and 'interrogate' them for evidence of causal significance. This trend is, as we have pointed out, part of a larger trend that many hope will eventually lead to the establishment of a genuinely *historical* social science.

Questions

1. In what ways do questions of timing and temporality matter to our understanding of specific political phenomena?
2. What are the two forms of temporality discussed in this chapter, and what are the methods that help to elucidate each of them?
3. Political researchers have become interested in the role narratives can play in theory-building. What methods can researchers use to convert a historical narrative into an *analytical* causal explanation?
4. What sorts of questions can the three types of comparative-historical analysis outlined in this chapter help you to answer?
5. What are some examples of research questions that process tracing can be used to address?
6. How do key concepts used by historical institutionalists help to explain institutional persistence and change?
7. In what ways might you address the issue of selection bias when using historical writing as data?

Guide to Further Reading

Beach, D. and R. B. Pedersen (2013), *Process Tracing Methods—Foundations and Guidelines* (Ann Arbor: University of Michigan Press).
This book identifies three different variants of process tracing that can be used to address different sorts of research questions. It also provides a detailed discussion of causal mechanisms and their role in process tracing, as well as the logic of different process tracing tests.

Bennett, A. and J. T. Checkel (eds) (2015), *Process Tracing: From Metaphor to Analytic Tool* (Cambridge: Cambridge University Press).
This book is a guide to designing and implementing research using process tracing. It clarifies core concepts and identifies standards for good applications of process tracing.

Elman, C. and M. F. Elman (eds) (2001), *Bridges and Boundaries: Historians, Political Science, and the Study of International Relations* (Cambridge: MA: MIT Press).
This edited volume contains the reflections of political scientists and historians about how their respective disciplines go about studying international events.

Fioretos, O., T. G. Falleti, and A. Sheingate (eds) (2016), *Oxford Handbook of Historical Institutionalism* (New York and London: Oxford University Press).
This book brings together articles by leading scholars which detail the contribution of historical institutionalism to an expanding array of topics in the study of comparative, American, European, and international politics.

Goodin, R. E. and C. Tilly (eds) (2008), *The Oxford Handbook of Contextual Political Analysis* (New York: Oxford University Press).

This is a valuable collection of articles that addresses why contextual factors matter in political analysis. Each of its sections provides a set of contributions devoted to explaining why a particular contextual factor matters, including why philosophy, psychology, ideas, culture, history, place, population, and technology matter.

Mahoney, J. and D. Rueschemeyer (eds) (2003), *Comparative Historical Analysis in the Social Sciences* (Cambridge: Cambridge University Press).
The book considers the extent to which work over the past several decades has contributed to the development of comparative historical analyses of revolution, the conditions of and for democracy, and social policy formation. It also addresses work that has contributed to our understanding of time and periodization, institutionalism, and network theory.

Pierson, P. (2004), *Politics in Time: History, Institutions, and Social Analysis* (Princeton: Princeton University Press).
This influential text seeks to show the importance of timing and sequence on the policy-making process, focusing on path dependency and positive feedback, the role of conjuncture in timing, and the long-term nature of social and political processes.

Thies, C. (2002) 'A Pragmatic Guide to Qualitative Historical Analysis in the Study of International Relations', *International Studies Perspectives* 3(4).
This discusses the use of primary and secondary source material as data or evidence. It explores potential problems encountered by researchers who conduct archival research or rely on secondary source material produced by historians, and suggests guidelines for minimizing selection and confirmation biases in the use of historical source materials.

Roberts, Geoffrey (2006) 'History, Theory and the Narrative Turn in International Relations', *Review of International Studies* 32(4): 703–14.
This is a discussion of the 'turn' to history and narrative in the study of international relations, and the contribution of narrative history to International Relations theory. He illustrates the role that narrative history could play in the development of International Relations theory by considering one particular theoretical issue: the agency and structure debate.

Trachtenberg, M. (2006), *The Craft of International History: A Guide to Method* (Princeton: Princeton University Press).
This 'nuts and bolts' guide to how to do historical work in international politics includes valuable advice on how to read secondary historical sources and where to find primary sources.

References

Acemoglu, D. and J. A. Robinson (2005), *Economic Origins of Dictatorship and Democracy* (Cambridge: Cambridge University Press).

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

Anheier, H. and H. Katz (2006), 'Learning from History? Comparative Historical Methods and Researching Global Civil Society', in M. Glasius, M. Kaldor, and H. Anheier (eds), *Global Civil Society 2005/6* (London: Sage).

Bengtsson, B. and H. Ruonavaara (2010), 'Comparative Process Tracing in Housing Studies'. Paper presented at conference: Comparative Housing Research—Approaches and Policy Challenges in a New International Era, Delft, The Netherlands, 24/25 March 2010.

Bennett, A. and J. T. Checkel (2014), '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), 3–37.

Bryant, J. M. (1994), *Evidence and Explanation in History and Sociology: Critical Reflections on Goldthorpe's Critique of Historical Sociology. The British Journal of Sociology* 45(1) (March): 3–19.

Centeno, M. A. and A. E. Ferraro (2013), *State and Nation Making in Latin America and Spain: Republics of the Possible* (Cambridge: Cambridge University Press).
Charrad, M. (2001), *States and Women's Rights: The Making of Postcolonial Tunisia, Algeria, and Morocco* (Berkeley, CA: University of California Press).
Cornell, S. E. and D. Hartmann (2007), *Ethnicity and Race: Making Identities in a Changing World*, 2nd edition (Thousand Oaks, CA: Pine Forge Press).
Daniel, G. R. (2006), *Race and Multiraciality in Brazil and the United States: Converging Paths?* (University Park, PA: Pennsylvania State University Press).
Dannreuther, R. (2010), 'Understanding the Middle East Peace Process: A Historical Institutionalist Approach', *European Journal of International Relations* 17(2): 187–208.
Deflem, M. and A. L. Dove (2013), 'Historical Research and Social Movements', in D. A. Snow, D. Della Porta, B. Klandermans, and D. McAdam (eds), *The Wiley-Blackwell Encyclopedia of Social and Political Movements* (Malden, MA: Wiley-Blackwell), 560–3.
De Fronzo, J. (2015), *Revolutions and Revolutionary Movements*, 5th edition (Boulder, CO: Westview).
Doornbos, M. (2006), *Global Forces and State Restructuring: Dynamics of State Formation and Collapse* (Basingstoke: Palgrave).
Elkins, Z., Ginsburg, T., and Melton, J. (2009), *The Endurance of National Constitutions* (Cambridge: Cambridge University Press).
Fioretos, O. (2011), 'Historical Institutionalism in International Relations', *International Organisation* 65(2) (April): 367–99.
Foran, J. (ed.) (2003), *The Future of Revolutions: Rethinking Radical Change in the Age of Globalization* (London: Zed Press).
Gandhi, J. (2008), *Political Institutions under Dictatorship* (New York: Cambridge University Press).
Gershenkron, A. (1962), *Economic Backwardness in Historical Perspective* (Cambridge, MA: Belknap Press of Harvard University Press).
Goemans, H. E. (2000), *War and Punishment: The Causes of War Termination and the First World War*. Princeton: Princeton University Press.
Goodwin, J. (2001), *No Other Way Out: States and Revolutionary Movements, 1945–1991* (Cambridge: Cambridge University Press).
Gough, I. and G. Wood (2004), *Insecurity and Welfare Regimes in Asia, Africa and Latin America* (Cambridge: Cambridge University Press).
Griffin, L. J. (2007), 'Historical Sociology, Narrative and Event-Structure Analysis: Fifteen Years Later', *Sociologica* 3/2007, doi: 10.2383/25956
Griffin, L. J. and R. R. Korstad (1998), 'Historical Inference and Event-Structure Analysis', *International Review of Social History* 43, Supplement: 145–65.
Haggard, S. and R. R. Kaufman (2008), *Development, Democracy, and Welfare States* (Princeton, NJ: Princeton University Press).
Hall, J. R. (2007), 'Historicity and Sociohistorical Research', in W. Outhwaite and S. P. Turner (eds), *The Sage Handbook of Social Science Methodology* (London: Sage Publications), 82–99.
Hechter, M. (2000), *Containing Nationalism* (Oxford: Oxford University Press).
Herbst, J. (2000), *States and Power in Africa: Comparative Lessons in Authority and Control* (Princeton, NJ: Princeton University Press).
Hibou, B. (2005), 'The Privatisation of the State: North Africa in Comparative Perspective', in K. Schlichte, (ed.), *The Dynamics of States: The Formation and Crises of State Domination* (Aldershot: Ashgate), 71–96.
Hooghe, M., T. Reeskens, D. Stolle, and A. Trappers (2009), 'Ethnic Diversity and Generalized Trust in Europe: A Cross-National Multilevel Study', *Comparative Political Studies* 42(2) (February): 198–223.
Htun, M. (2003), *Sex and the State: Abortion, Divorce, and the Family under Latin American Dictatorships and Democracies* (Cambridge: Cambridge University Press).
Huber, E. and J. D. Stephens (2001), *Development and Crisis of the Welfare State: Parties and Policies in Global Markets* (Chicago: University of Chicago Press).
Hui, V. T. (2005), *War and State Formation in Ancient China and Early Modern Europe* (New York: Cambridge University Press).
Inglehart, R. and P. Norris (2003), *Rising Tide: Gender Equality and Cultural Change Around the World* (Cambridge: Cambridge University Press).
Knight, E. and J. Knox-Hayes (2015), 'Creating Legitimate Authority for Environmental Governance and New Market Creation: A Case Study from Australia', *Competition & Change* 19(1): 36–55.
Kohli, A. (2004), *State-Directed Development: Political Power and Industrialization in the Global Periphery* (New York: Cambridge University Press).
Little, D. (2000), 'Explaining Large-Scale Historical Change', *Philosophy of the Social Sciences* 30(1) (March): 89–112.
Liu, J. H. and D. J. Hilton (2005), 'How the Past Weighs on the Present: Social Representations of History and their Impact on Identity Politics', *British Journal of Social Psychology* 44: 537–56.
Liu, J. H., C. G. Sibley, and L.-L. Huang (2014), 'History Matters: Effects of Culture-specific Symbols on Political Attitudes and Intergroup Relations', *Political Psychology* 35(1) (February): 57–79.

Lustick, I. (1996), 'History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias', *American Political Science Review* (September): 605–18.
Mahoney, J. (2010), *Colonialism and Postcolonial Development* (Cambridge: Cambridge University Press).
Mahoney, J. and D. Rueschemeyer (2003), 'Comparative Historical Analysis: Achievements and Agendas', in J. Mahoney and D. Rueschemeyer (eds), *Comparative Historical Analysis in the Social Sciences* (Cambridge: Cambridge University Press), 3–40.
March, J. G. and J. P. Olsen (1984), 'The New Institutionalism: Organizational Factors in Political Life', *The American Political Science Review* 78(3) (September): 734–49.
McMichael, P. (2012), *Development and Social Change: A Global Perspective*, 5th edition (London: Sage).
Oberschall, A. (2007), *Conflict and Peace Building in Divided Societies: Responses to Ethnic Violence* (London: Routledge).
O'Connor, J. S. (2015), 'The State and Gender Equality: From Patriarchal to Women-Friendly?', in S. Leibfried, F. Nullmeier, E. Huber, M. Lange, J. Levy, and J. Stephens (eds), *The Oxford Handbook of Transformations of the State* (New York: Oxford University Press), 482–98.
Parsa, M. (2000), *States, Ideologies, and Social Revolutions: A Comparative Analysis of Iran, Nicaragua, and the Philippines* (Cambridge: Cambridge University Press).
Pierson, P. (2004), *Politics in Time: History, Institutions, and Social Analysis* (Princeton, NJ: Princeton University Press).
Pierson, P. and T. Skocpol (2002), 'Historical Institutionalism in Contemporary Political Science', in I. Katznelson and H. V. Milner (eds), *Political Science: State of the Discipline* (New York: W. W. Norton), 693–721.
Pomeranz, K. (2001), *The Great Divergence: China, Europe, and the Making of the Modern World Economy*, revised edition (Princeton, NJ: Princeton University Press).
Richards, M. D. (2009), *Revolutions in World History* (London: Routledge).
Robinson, J. (2010), 'Elites and Institutional Persistence'. WIDER Working Paper 2010/085 (Helsinki: UNU-WIDER).
Rodrik, D. (2007), *One Economics, Many Recipes* (Princeton, NJ: Princeton University Press).
Salcedo, C. and M. E. Akoorie (2013), 'Foreign Direct Investment in Chile: Historical Process, Changing Political Ideologies and the Responses of MNEs', *AD-minister* 23: 103–29.
Skocpol, T. (1979), *States and Social Revolutions: A Comparative Analysis of France, Russia, and China* (Cambridge, MA: Cambridge University Press).
Skocpol, T. and M. Somers (1980), 'The Uses of Comparative History in Macrosocial Inquiry', *Comparative Studies in Society and History* 2(2) (April): 174–97.
Swank, D. (2002), *Global Capital, Political Institutions, and Policy Change in Developed Welfare States* (New York: Cambridge University Press).
Tilly, C. (2003), *The Politics of Collective Violence* (Cambridge: Cambridge University Press).
Van Evera, S. (1997), *Guide to Methods for Students of Political Science* (Ithaca, NY: Cornell University Press).
Waldner, D. and B. Smith (2015), 'Rentier States and State Transformations', in S. Leibfried, F. Nullmeier, E. Huber, M. Lange, J. Levy, and J. Stephens (eds), *The Oxford Handbook of Transformations of the State* (New York: Oxford University Press), 714–29.
Walzer, M. (2015), *The Paradox of Liberation: Secular Revolutions and Religious Counterrevolutions* (New Haven, CT: Yale University Press).
Wood, E. J. (2000), *Forging Democracy from Below: Insurgent Transitions in South Africa and Brazil* (New York: Cambridge University Press).

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

Endnote

1. The program is freely available for online or offline use (at http://www.indiana.edu/~socpsy/ESA). The website also includes useful tutorials and program manual files.

11 Surveys

Chapter Summary

This chapter examines the principles of survey research and discusses the issues and problems associated with different stages of the design process. In doing so, it examines questionnaire design, sample design, and interview techniques and discusses the common sources of error that affect survey research and what can be done to try and avoid or minimize them. The chapter examines:

- questionnaire design;
- measurement error;
- sampling design;
- sampling error;
- interview mode.

Introduction

This chapter examines the principles of survey research, and discusses the issues and challenges involved in carrying out a good survey. Despite a number of weaknesses, surveys continue to flourish and are frequently used in political research to investigate a wide range of political phenomena. They combine two things: the ancient but extremely efficient method of obtaining information from people by asking questions; and the modern random sampling procedures that allow a relatively small number of people to represent a much larger population (Schuman and Presser 1996: 1). As such, they are a valuable resource for examining a wide range of topics, and when done well, provide an accurate and reliable insight into what ordinary people think about politics and how they participate in politics.

A successful survey involves many different steps. Before thinking about questionnaires or samples, however, it is first necessary to establish what the main aims of the research project are, and to clarify the key concepts, issues, and theory that you want to examine (see Chapters 4 and 5). The survey will only give you the answers to the questions that are asked,

and it is therefore important to have a clear idea about what it is that you want to find out before the survey is carried out, since it will be too late afterwards if a new idea comes to you when you are doing the analysis. The decisions that are made at this stage will influence each of the subsequent steps that are taken.

The next step—which is the focus of this chapter—is to decide how the questionnaire is going to be designed and how the sample is going to be selected. Surveys are about asking questions, and the type of question you ask can affect the sort of answer you get in three main ways. What question do you ask? Who do you ask the question to? And how do you ask the question? The purpose of this chapter is to examine each of these in turn. The first part of the chapter deals with *what* you ask and examines issues relating to questionnaire design. The second part deals with *who* you ask and covers sampling design, and the third part deals with *how* you ask questions, and looks at different types of interview technique.

The elements of a good survey

Surveys have both a descriptive purpose and an analytical purpose. The descriptive purpose is to try and provide accurate information about what people think and do in a given population; that is, to try and understand what happens. This is a valuable end in itself, and whether the survey is designed to produce population estimates about income or inequality, trust in parliament, or satisfaction with democracy, or how many people think terrorism is a legitimate form of political protest, the answers to these sorts of questions can be extremely revealing. However, surveys can be used for much more than just this. The analytical purpose is to provide data that can be used for theory testing or hypothesis testing to try and understand why things happen. It is one thing to know that a certain percentage of the population has lost faith in democracy; it is quite another to know the reasons why. Surveys are a powerful tool for trying to answer these questions. The researcher develops theories or explanations about why things happen, and uses surveys to collect data to test those explanations against what is observed in the real world. If the expectations are correct, then this lends support to the theory. If the expectations are wrong, there are two main possibilities. Either the theory is wrong, in which case it is back to the drawing board to try and come up with something more plausible, or the measurement or testing of the theory is wrong, and is somehow flawed. In this case, there may be nothing wrong with the theory or explanation per se, but the way in which key concepts have been operationalized (see Chapter 7) is bad. Obviously, we want to try and minimize the possibility of the latter occurring, and this means we must be very careful how we collect the information that we need, and how we then analyse it.

The focus of this chapter is on how surveys can be used to collect information. Chapters 15 to 17 deal with issues of analysis, but before we reach that stage it is necessary to be confident that the data we wish to analyse are reasonably sound. One of the great strengths of survey research is that it helps us to make general claims about what different sections of society or different sub-groups of the population actually think and do. It thus gives voice to people that might not otherwise be heard. But a major weakness is that surveys can, and frequently do, misrepresent what people think and do, and thus create misleading information.

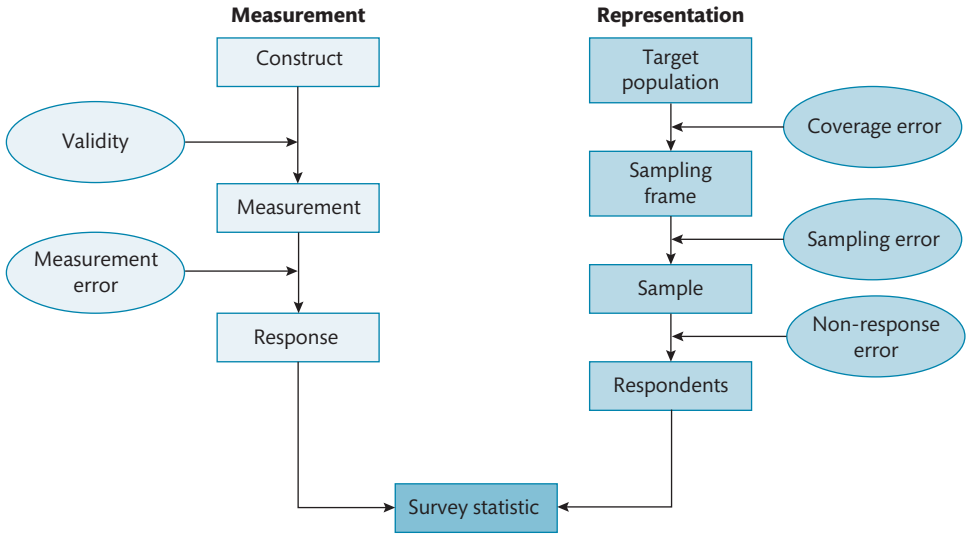


Figure 11.1 Sources of error in survey research

The extent to which surveys misrepresent the 'real' or 'true' attitudes and behaviour of the people they seek to study can be thought of as error. The purpose of a good survey is to try and minimize this error. In practice, it is rarely possible to eliminate all error (indeed, it is virtually impossible), but with good research design and careful administration, it is possible to reduce error, and perhaps more importantly, to be aware of what error remains.

This chapter focuses on survey error. It describes the main types of error that are found in surveys, and where possible provides information about what to do in order to try and reduce them. There are two important sources of error that we consider. These are highlighted in Figure 11.1 (see also Groves et al. 2004). The first is to do with measurement error, and refers to the ways in which surveys use questions to try and measure different social and political phenomena, such as political attitudes, opinions, and behaviour. The second is to do with sampling error, and refers to the ways in which respondents are chosen or selected to complete the survey and the implications this has for the representativeness of the sample.

What you ask: questionnaire design

Questionnaire design is a critical component of survey research. If questions are badly designed, then the results they produce will be flawed, no matter how large the sample size or how sophisticated the statistical analysis. Badly phrased questions or poorly measured concepts can render the research exercise useless before it has even begun and waste a lot of money in the process. It is therefore critical to make sure the **questionnaire** does what it's supposed to do as well as it can.

Validity: from concepts to survey questions

One of the most common uses of surveys is to measure the attitudes and behaviour of ordinary people. Many of these phenomena—or constructs—that we are interested in measuring are not directly observable. We therefore have to ask questions to try and measure the constructs we are interested in. Generally, it is thought that it is easier to measure behaviour, since this is more concrete. Past behaviour is in a sense a 'fact'; someone either voted or they didn't, they are either a member of a political party or they are not, and so on. By contrast, questions relating to attitudes and opinions are more abstract, and so more difficult to pin down and measure through the use of one or more questions. Although there is some truth in this, in reality there are difficulties associated with measuring both attitudes and behaviour. The harder a concept is to define, the harder it is to measure, so the first step is to clarify exactly what it is that we seek to examine (see the discussion of conceptualization and operationalization in Chapter 7). Concepts that are vague or multifaceted need to be pinned down and perhaps broken up into smaller parts.

Once we have a definition that we are happy with, the next step is to think about how it can be measured. This involves moving from the abstract to the specific. This is one of the hardest jobs in survey research, and it can easily go awry. What question (or questions) taps into the general idea that you are interested in exploring? These questions—or indicators—are the bread and butter of survey research. Whether we are measuring behaviour or attitudes, our primary concern is that the question we use is both valid and reliable. Validity means that the question measures what we think it does and reflects the concept that we are trying to address. It is an issue of how the question is used and the purpose it is used for. For example, if we are interested in a general concept, such as social status, we might use a question on educational level to try and measure this (de Vaus 2002: 53). The issue of validity is not whether we have measured education properly, but whether this is a suitable measure of our concept—social status. This is a frequent problem in social and political research. Often questions are used in ways for which they are not ideally suited. There may therefore not be anything wrong with the question per se, but the purpose that it is used for may not be appropriate. The validity of a measure therefore depends upon how we define the concept. This is always going to be open to debate, though is obviously more straightforward in some cases than others.

There are no easy ways to resolve this. But there are four basic ways in which we can try and assess validity (see Moser and Kalton 1971 and de Vaus 2002 for an extended discussion). The first and most straightforward is face validity. Face validity simply means: on the face of it, does the question intuitively seem like a good measure of the concept? If the answer is no, then we definitely have a problem, but what is intuitively yes for one person may not be so for another. Content validity examines the extent to which the question covers the full range of the concept, covering each of its different aspects. **Criterion validity** examines how well the new measure of the concept relates to existing measures of the concept, or related concepts. **Construct validity** examines how well the measure conforms to our theoretical expectations, by examining the extent to which it is associated with other theoretically relevant factors. Ultimately, though, the matter of validity is one of judgement.

Reliability means the question is answered in the same way on different occasions. It is clear that an unreliable question lacks validity. But a reliable question need not be valid, because it could be measuring something other than what it was designed to measure, and reliability should not compensate for low validity.

Measurement error: developing indicators

A well-constructed question may or may not be valid. However, a badly worded question is never valid because it will not be a good measurement instrument. The next important issue to address then, as shown in Figure 11.1, is measurement error. By measurement error or error in observation, we mean the difference between what we record as the respondent's answer to a question and what is the respondent's real or true answer to that question.

On the whole, respondents will try to answer a question truthfully and to the best of their ability, but they may still not provide an accurate answer to the question they are asked for a number of reasons. It may be that they simply have not understood the question (which refers to problems of comprehension) or they cannot remember the answer (problems of recall). However, there are also more subtle reasons why they may not provide a 'true' answer. One reason is to do with the suggestibility of respondents. Some people may not be sure how to answer the question because they have never really thought about it before, and, rather than admit as much, they look for cues in the way in which the question is asked to give them clues how to answer. Another reason is that they may hold unfashionable or even unsavoury views, and thus be unwilling to reveal them. The following section discusses each of these potential sources of measurement error in turn.

Question wording comprehension

The literature on the wording of questions is vast (see Moser and Kalton 1971 and Tourangeau, Rips, and Rasinski 2000 for good overviews). Respondents will interpret the question as best they can; but the more difficult a question is to understand, the more likely it is that respondents will interpret the question in different ways, introducing confusion about what exactly is being measured. We shall confine ourselves to some aspects of wording that are of general importance in survey research.

A common error is to ask a general question when a more specific question is needed. For example, if one is interested in the extent to which the general public trust their political representatives, it might be tempting to ask the following question: 'Do you trust politicians?' However, this question is unsatisfactory. It is not clear what trust means in this context. The respondent might trust politicians not to take bribes, but may not trust them to tell the whole truth when they are in a tight corner. The question fails to provide the respondent with the necessary frame of reference to give a meaningful answer. There may, of course, be a number of distinct frames of reference. One might be to tell the truth, one might be to put the interests of the country first, but whatever they are, they should be made explicit.

Ambiguity means that the question can be interpreted in different ways, and so in effect respondents may be answering different questions. A question can be ambiguous for a wide

variety of reasons, and it may not always be obvious that it is ambiguous until you see how it is interpreted by different people. That is why it is so important to pre-test questions by carrying out exploratory research and piloting the survey (see Chapter 12 on interviews as a way of doing this). One particular type of ambiguity arises with **double-barrelled questions**. This is when the question incorporates two distinct questions within one. So, for example, consider the question: 'Do you agree or disagree with the following statement: "Tony Blair lied to parliament about Iraq and should be tried for war crimes."' You might think he lied but should not be tried for war crimes, and you might think he didn't lie but should be tried for war crimes, so how do you answer? Better to break the question up into two separate questions and ask about whether or not you thought he lied in parliament, and whether or not he should be tried for war crimes separately.

Related to the double-barrelled question is the question that relies on **faulty presupposition**. This means that the question assumes something that is not necessarily true. Moser and Kalton refer to these types of questions as **presuming questions**. For example, consider the following statement: 'Democracy is the best form of government because it makes leaders accountable.' This question rests on the premise that democracy is indeed the best form of government. In this context, it is hard to know what it means if a respondent disagrees with the statement. It could mean, 'I disagree; democracy is not the best form of government', or it could mean 'I disagree; accountability is not the main reason why democracy is the best form of government.' Likewise, the question 'Who did you vote for in the last General Election?' presupposes that the respondent voted. These types of question are best asked only after a **filter** question has revealed that the respondent did vote in the last election, and so, once again, are best asked separately.

The next three problems involve the meaning of words or phrases in the question. It is important that the question has the same meaning for different respondents. Vague concepts or vague quantifiers should therefore be made as concrete as possible. Vague questions encourage vague answers. If you ask someone whether they talk about politics regularly or occasionally, the meaning of their answer will be vague. As Moser and Kalton observe, this common choice of alternatives is strictly illogical since the word 'occasional' refers to frequency, whereas the word 'regular' does not. Many everyday terms are vague, and respondents may interpret them in different ways. This is a particular issue in comparative surveys, where the same term or word can have different connotations in different contexts. For example, the word 'nationalist' might have positive connotations in some countries, but negative connotations in other countries, particularly when 'nationalist' movements have been linked to acts of violence or repression. It is therefore important to investigate whether the question has the same (equivalent) meaning in the different countries where it is being asked. This is known as establishing equivalence of meaning.

Similarly, unfamiliar terms or jargon should be avoided. The aim in a question wording is to try and communicate with respondents as clearly as possible in a language they will understand. People who write surveys are often experts in the subject, and they may overestimate how familiar the respondents are likely to be with the terminology that they use. So, even if it is perfectly clear to the question designer what is meant by fiscal policy, democratic deficit, humanitarian intervention, or bilateral trade agreement, it might not be for the person answering the question. Moser and Kalton (1971) counsel question designers to put themselves in the position of their typical—or least educated—respondent. Many

words which are common to university-trained survey experts are far from common in ordinary conversation. Even words like hypothetical, aggravate, deprecate, and hundreds more are in this category, and Gowers (1954) mentions many words which can often be replaced by simpler alternatives, such as 'help' instead of 'assist'; 'think' instead of 'consider'; 'begin' instead of 'initiate'; 'country' instead of 'state'; and 'enough' instead of 'sufficient'. For surveys of the general population (rather than elites, which is discussed in Chapter 12), Moser and Kalton (1971: 321) advise that the first principle in wording is that questions should use the simplest words that will convey the exact meaning, and that the phrasing should be as simple and informal as possible. The more questions sound like an ordinary (polite) conversation, the smoother the interview will be.

To ensure that respondents all understand the question and interpret it in the same way, it is therefore recommended to (1) define terms carefully; (2) avoid unfamiliar, ambiguous, or imprecise words; (3) avoid complicated wording / too many clauses; (4) avoid two or three questions in one. To help overcome comprehension problems, piloting the questionnaire is crucial. Observing how respondents answer questions can give a clear indication of how well they have understood them. Do they ask for the question to be repeated? Do they ask for clarification? Do they stare blankly at the interviewer?

Recall problems

However, even when respondents understand the question, measurement error can still arise in a variety of ways. Another potential source of error in survey responses is to do with memory failure—or **recall problems**. This is particularly relevant for questions about past behaviour or past actions that the respondent may have undertaken, such as whether they voted in the last election, and, if so, for which party (see Heath and Johns 2010). Sometimes respondents cannot remember the relevant events at all, and sometimes they partially remember or inaccurately remember. Wright (1993) suggests that one reason for this is that less politically engaged respondents are less likely to retain the information of who they voted for in their memory. Voters simply forget which party or candidate they voted for. Respondents struggling to remember how they voted are probably more likely to infer that they voted for the winner (Wright 1993; Atkeson 1999). This gives rise to the so-called **bandwagon** effect.

In addition, another type of memory-related problem is to do with what Belli et al. (1999) call source-monitoring error. This problem can affect questions about all sorts of behaviour, from turnout to protest to acts of aggression, particularly if the question is concerned with behaviour within a specific time frame, such as the past twelve months. Respondents can generally remember that they have done such an activity, but cannot recall accurately when it was, and so they confuse real participation within the specified time frame with earlier experiences. This is also known as **telescoping**. As more time passes since the event, we make more errors in dating when it took place.

There are a number of strategies that can be used to try and deal with these problems. Perhaps the most sensible is not to put the respondents' memory under too much strain in the first place by asking them to recall events in the distant past, or to recount the number of times they have done something in a fairly lengthy time period. For this reason, questions tend to focus on events that have taken place in the past seven days for

relatively common activities, such as reading the newspaper, or the past twelve months for less common activities, such as taking part in a protest.

Deliberate misreporting

However, even if respondents understand the question and can remember the answer, they may not always answer entirely truthfully. It is desirable that respondents answer the question. Non-response (or **item non-response** as it is known when it refers to a specific question) can introduce error. Refusals may happen for a variety of reasons, but are more common when the question deals with sensitive issues. Respondents may not want to declare how much money they earn, or which political party they support (particularly in polarized political contexts). For these reasons, sensitive questions tend to go at the end of questionnaires, although this needs to be balanced against issues relating to question order effects.

Respondents generally want to be helpful, but at the same time there is a tendency, which has been found in many studies, that they also want to present themselves in a good light. This is known as **social desirability bias**, in which respondents over-report socially 'desirable' behaviour and under-report 'undesirable' behaviour. At one extreme, respondents may be unwilling to admit that they hold racist views, or vote for extreme-right parties, and so instead will opt for another answer (perhaps saying they don't know). But even more moderate views and forms of behaviour are susceptible to the same biases. For example, surveys frequently find that respondents over-report the true extent to which they have voted. This in part may stem from a desire to appear a good citizen and to have fulfilled one's civic obligations (though the extent to which voting is considered a civic obligation obviously varies across contexts). One way to counteract this relatively mild social desirability bias is to load the question in a certain way that gives the respondent a ready-made excuse for not having done what they may think is expected of them. Loading a question means wording it in a way that invites a particular response, as discussed later with reference to leading questions. In this case, the question is loaded to try and invite the socially undesirable response. Sudman and Bradburn (1982) distinguish several strategies for doing this: ranging from the 'everybody-does-it' approach to the 'reasons for doing it approach'. Questions on voting generally borrow from these strategies, and ask something like this, which is taken from the British Election Study (BES):

> Talking with people about the general election on May 5th, we have found that a lot of people didn't manage to vote. How about you, did you manage to vote in the general election?

These strategies help to reduce the apparent sensitivity of the question. Similarly, embedding one sensitive question (e.g. on vandalism) among other more sensitive questions (e.g. on terrorism) may make the sensitive item of interest seem less threatening by comparison.

Question order effects

So far, the problems discussed have been relatively straightforward, but there is another set of measurement problems that is less easy to define. One of the major problems that affect survey research is to do with the order in which questions are asked—**question order**

effects. Apart from sampling error, question order effects are probably the most commonly cited explanation for an unexpected or anomalous survey finding (Schuman and Presser 1996: 24). The context in which a question is asked or its position in relation to other questions can have a bearing on the answers that are given to it. The most obvious manifestation of this is when respondents try to appear consistent by answering one question to fit in with how they answered a previous question. To show the sort of effect that the order of questions can have, we draw on a survey experiment carried out by Schuman and Presser (1996: chapter 2), in which alternative versions of a questionnaire were asked to random halves of the sample. Respondents were asked two questions, about communist journalists and American journalists. However, the order in which they were asked these questions varied across the two half-samples. Survey experiments of this type are frequently used in political research for a wide variety of purposes (see Heath et al. 2015).

> Do you think the United States should let Communist newspaper reporters from other countries come in here and send back to their papers the news as they see it?
>
> Do you think a Communist country like Russia should let American newspaper reporters come in and send back to America the news as they see it?

From Table 11.1, we can clearly see that respondents were much more likely to say that they would allow communist reporters into the United States after they had answered the question on the American reporter first. When the question on the communist reporter was asked first, only 54.7 per cent of the respondents thought communists should be let into the United States, but when the American question was asked first, this percentage increased to 63.7 per cent (a statistically significant increase). Similarly, respondents were less likely to say they would allow American reporters in to communist countries after they had answered the communist question first, with the percentage dropping from 81.9 per cent to 74.6 per cent (also a statistically significant change). Schuman and Presser interpret this to suggest that respondents answer the question in terms of pro-American and anti-communist sentiments, whichever question is asked first. But having answered the first question, they follow a norm of reciprocity, in which a substantial number feel compelled to provide an answer that is consistent with their previous response.

This type of problem can affect a wide range of questions, and it is not obvious how to deal with it. Generally, it is thought that it is preferable to ask the more important questions first, so they are not contaminated by responses to less important questions. For this reason, in most National Election Studies, questions on which party respondents voted for generally

Table 11.1 Order effects on communist and American newspaper reporter items in 1980

Communist reporter first		American reporter first	
Yes to communist reporter	Yes to American reporter	Yes to communist reporter	Yes to American reporter
54.7%	74.6%	63.7%	81.9%

Source: Adapted from Schuman and Presser (1996).

tend to come at the beginning of the questionnaire. However, the downside of this is that it risks contaminating subsequent questions that are aimed at measuring why people voted as they did. To try and reduce the possibility of this happening, it is perhaps a good strategy to leave gaps between questions where this type of contamination is thought to be most likely.

Question wording effects

Another group of problems arise from what can broadly be termed **question wording effects**. Respondents may be more (or less) likely to report certain types of behaviour or attitude, depending upon the way in which a question is phrased or the answer is recorded. A good example of this is the difference between **open-ended** and **closed questions**, and how response format can influence the answers that are given to the same question. A 'closed' question provides the respondent with a set of answers to the question and the respondent must select the response that reflects his view most closely. By contrast, an open-ended question does not provide the respondent with any answer-option to choose from, and the respondent must formulate his own response himself, which the interviewer will then record verbatim.

There are pros and cons attached to each format. A major advantage of open-ended questions is that they give the respondent the opportunity to voice exactly what he thinks. They allow for a detailed and reasoned response, which can reveal the logic and rationale behind a respondent's answer. Answers are not prompted, and so reveal what the respondent actually thinks. However, there is also a number of disadvantages of using open-ended questions. They may provide a lot of information, but this information then needs to be coded in order to make sense of the answers. This can be very time-consuming. Many responses may be irrelevant or off-topic, and some groups of people may be more likely to respond than others, particularly those who are more educated or more interested in the topic.

For these reasons, closed questions are the standard procedure for most questions. Attitude questions almost always use a closed format, with the answer categories forming a scale. Closed questions provide the respondent with a pre-defined set of response alternatives. They are quick and easy for the respondent to answer and for the interviewer to code, and do not discriminate against the less talkative or eloquent respondents. They can also remind respondents of answers that they might not otherwise have retrieved from their memory (Schuman, Ludwig, and Krosnick 1986). However, this kind of prompting may be used by respondents to save themselves the mental effort of producing a valid answer themselves. Moreover, it is quite likely that the predetermined list of alternatives will exclude some issues that are genuinely salient to some people. This is a problem given survey respondents' general reluctance to insist on a response that is not explicitly offered (and thus legitimized) by the question.

These points are well-illustrated by survey experiments comparing open and closed versions of questions on what respondents consider to be the most important political issue facing the country (Schuman and Presser 1996). Eight alternatives were offered in the closed format, though respondents did have the option of 'Other'. The distribution of responses in the two versions is shown in Table 11.2.

The first point to note is that, when unprompted, 36.5 per cent of respondents named one of a wide variety of problems that were not included in the list of eight alternatives.

Table 11.2 Comparing open and closed questions measuring the most important issue facing the country

	Closed	Open
	Which of these is the most important problem facing the country at present?	What do you think is the most important problem facing the country at present?
Food and energy shortages	6.2	1.7
Crime and violence	36.1	16.1
Inflation	13.0	13.6
Unemployment	20.4	19.6
Decreased trust in government	10.2	3.1
Busing	1.1	1.1
Breakdowns of morals and religion	9.5	5.8
Racial problems	1.7	2.5
Other issues/problems	1.9	36.5
Total	100.0	100.0
N	422	449

Source: Adapted from Schuman and Presser (1996).

In the closed version, hardly any respondents insisted on these 'other' issues; instead they chose from among the options explicitly offered. However, and this is the second key point, they did not choose between those eight issues in the same proportions as respondents to the open-ended version. Instead, they opted in disproportionately large numbers for 'crime and violence' and 'decreasing trust in government'. In contrast, the percentages for 'inflation' and 'unemployment' are barely different across the two versions. It seems that economic issues come to respondents' minds with or without prompting, but that crime is a problem the importance of which respondents need a reminder. (See also Heath and Johns 2010.)

It is not straightforward to assess which of these questions produces the truer picture of issue salience. It may be that respondents perceive crime to be a local issue, and so do not consider it when asked the open-ended version of the questionnaire. Thus the open question provides the narrower frame of reference. Which format is preferred is therefore difficult to resolve. The argument for closed questions is stronger in those cases where survey designers are confident that their list would cover most, if not all, of the more common responses. The best route to such confidence is to use open questions in a pilot survey, and then to base the closed list on the results from that pre-testing.

A further example of question wording effects relates to the tone of the wording. This is illustrated by other split ballot experiments that examine support for freedom of speech in the United States (Schuman and Presser 1996). The results from these experiments show that when respondents were asked if the government should *forbid* speeches against democracy, support for freedom of expression was much higher than when respondents were asked if the government should *allow* speeches against democracy.

The reason why logically equivalent questions can produce different responses according to the tone of wording that is used can occur for a variety of reasons. In general, those

who are most susceptible to wording effects are probably those for whom the question does not have great resonance or importance. People who design questionnaires for political research tend to be very interested in the subject, and to have thought a great deal about the issues. However, these people also tend to be outliers, and, for many respondents, the subject matter may not be of any great personal importance. These respondents with weaker attitudes may therefore be responsive to a variety of external influences, ranging from a story on the previous night's news to a nuance in the question wording (Zaller 1992; Tourangeau, Rips, and Rasinski 2000).

Balanced questions

A similar way in which question wording can influence responses is to do with whether it is balanced or not. In general, questions should be neutral. In practice, this is difficult to achieve, and questions may vary considerably in terms of how neutral or balanced they are. One example of an unbalanced question is what is often described as a **leading question**. A leading question indicates to respondents that certain responses are more useful or acceptable to those asking the question (Belson 1981). Generally, they are to be avoided, since they can prompt the respondent to answer the question in a certain way. One famous example of this type of question that is still widely used is an item designed to measure whether respondents identify with a political party.

The original question developed for the National Election Studies in the United States is: 'Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or what?' This question has been exported to many other contexts, including the BES, which asks: 'Generally speaking, do you think of yourself as Labour, Conservative, Liberal Democrat, or what?'

The use of 'or what' at the end might be considered somewhat rude, even threatening, and could be taken to imply that not having such an identification is a bit odd. Respondents may therefore feel pressured to say that they do identify with one of the parties. Similarly, mentioning only the main parties may make respondents less likely to say that they think of themselves as partisans of other parties which are not mentioned. Certainly, alternative forms of asking this question in an open-ended format reveal much lower levels of identification (Heath and Johns 2010).

But questions can also be unbalanced in other ways. And a perfectly balanced question is difficult to achieve. **Acquiescence bias** refers to a tendency among respondents (or some types of respondent) to agree with attitude statements presented to them. There are several different interpretations for this. Implicit in many of these interpretations is that some groups of people, such as the poorly educated, are more likely to lack true attitudes on issues presented to them. Converse (1964) labelled these non-attitudes. Since they have no real answer to the question, they tend to just agree. Acquiescence bias is therefore likely to be more pronounced on issues or attitudes that are of low salience to the respondent, in which he lacks clearly defined views, or in cases where the question is not clearly understood. The most obvious remedy to this problem is to mix pro and anti statements, so that at least the extent of the problem can be identified. Related to acquiescence bias are something known as recency effects, which refer to the order in which responses are presented to the respondent in closed-format questions. Recency effects are said to occur when respondents select the last option

they are offered. They are generally thought to be more likely to occur when the question is long or complicated, or when the respondent does not have a 'crystallized' opinion.

Response format

Another important aspect of questionnaire design relates to how to record the respondents' answers, or more specifically what types of response option to provide them with. There are many issues at stake here. The first concerns the number of response categories that are provided—and whether dichotomous (yes/no) or polytomous (e.g. levels of agreement) response categories are used. If polytomous response scales are to be used, then how many should be used and should a mid-point be included? There is little standard practice and there is considerable variation in terms of what is generally used—though, for most topics, five categories are probably sufficient.

Standard questions generally do not offer a 'Don't know' category, but will accept don't know as a valid response if it is spontaneously offered by the respondent. The rationale behind this approach is to try and probe an opinion. But in some cases this probing can invent data, as the respondents will be forced into providing an opinion that they do not really hold. Indeed, Converse (1964) suggests that much of the error that occurs in surveys comes from the random responses that are provided by people with no real opinion. To distinguish between 'real attitudes' and 'non-attitudes' it is therefore often a good idea to provide a question to gauge the respondents' level of interest or knowledge about the topic before (or after) asking about their opinion.

Who you ask: sample design

So far, we have discussed issues relating to questionnaire design, and how the questions you ask and the way in which they are asked can affect the answers you get. Equally important is the issue of sampling design. This is because surveys rely on samples, which are selected from the populations that we seek to study. In sampling terminology, the population of interest is a particular group of people or other units (such as households, businesses, schools, etc.) that the researcher wishes to study. Thus, for a study on voting behaviour in a given country, the population of interest may be all of the eligible voters. We rarely interview everybody in the population (when we do, it's called a census) because it is incredibly time-consuming and expensive, and so we must therefore choose who we select to interview. The purpose of the sample survey is to allow the researcher to make a generalization about the attributes of a given population without having to actually interview everyone in that population to find out what they think and do. In order to do this, the survey relies upon interviewing a sample of the population. How we do this has a big impact on the results we get. If the sample is chosen in a haphazard or subjective way, then there is little hope of making accurate generalizations about the wider population we are interested in. There are thus three main sources of error that can influence how representative the sample is of the population (see Figure 11.1). These relate to **coverage error** (what we choose our sample from), sampling error (how we choose our sample), and **non-response** (how many people from our sample complete the survey).

Defining the population: sampling frame

The *sample frame* is the set of the target population that has a chance to be selected in the survey sample. Once we have defined our population of interest, we must obtain a **sampling frame**, which contains a complete (or as complete as possible) list of all the population units. A frequently used sampling frame for national surveys is the Post Office Address file, which contains an almost complete list of addresses in the UK. It is important to remember that a random probability sample alone does not guarantee that a given sample will be representative if the sampling frame from which the sample is drawn does not cover the entire population. Under-coverage or coverage error can become a problem if certain segments of the population are left off the sampling frame. For example, the Electoral Register is one possible sampling frame for registered voters in the UK, but it will not tell us about people who are not registered to vote. It is therefore not such a good sampling frame for the population as a whole.

Choosing the sample: random probability sampling

A good sample is representative of the population from which it is drawn. This allows us to generalize our findings from the survey to the population. There are many different methods for drawing a sample. Since there is no way of knowing what our population looks like in every respect, it is difficult to assess how well we have matched our sample to the population. But there are some ways that will tend to produce better results than others. There are two main ways of drawing a sample: probability samples, which are drawn at random; and non-probability samples, which are selected more subjectively. To be representative in a sense appropriate for statistical inference, a sample must be a probability sample, where everyone in the population has a known, non-zero probability of being sampled. Generally, most academic surveys follow this approach. Non-probability samples are more frequently used in Market Research and 'Opinion Polls', and although they are not reliable in a statistical sense, they do often produce results that are reasonably accurate. However, when they do go wrong, they can do so spectacularly. (See Box 11.1.)

BOX 11.1 Sampling Design: How Not to Do It

A famous example of the dangers of non-probability sampling is the survey by the *Literary Digest* magazine to predict the results of the 1936 US presidential election. The magazine sent out about ten million questionnaires on postcards to potential respondents, and based its conclusions on those that were returned. This introduced biases in at least two ways. First, the list of those who were sent the questionnaire was based on registers such as the subscribers to the magazine, and of people with telephones, cars, and various club memberships. In 1936, these were mainly wealthier people who were more likely to be Republican voters, and the typically poorer people not on the source lists had no chance of being included. Second, only about 25 per cent of the questionnaires were actually returned, effectively rendering the sample into a volunteer sample. The magazine predicted that the Republican candidate Alf Landon would receive 57 per cent of the vote, when in fact his Democratic opponent, F. D. Roosevelt, gained an overwhelming victory with 62 per cent of the vote. The outcome of the election was predicted correctly by a much smaller probability sample collected by George Gallup.

Probability samples assign each unit in the sampling frame a known non-zero chance of being selected. The important thing is that the probability of being selected is known for each unit; it does not have to be equal. For example, survey researchers frequently over-sample small but interesting groups in the population so they can be analysed in greater detail. There are four main types of sample design based on probability sampling. These are: (1) **Simple random sampling** (SRS); (2) systematic sampling; (3) stratified sampling; and (4) cluster sampling.

SRS is perhaps the most straightforward type of probability sampling. Every unit in the sampling frame has the *same* probability of being selected. The units are assigned numbers, and then random number tables are used to select the sample. This method, though, is not without drawbacks. The first is that it can be very expensive to carry out when investigators have to interview the respondents personally because they may need to travel long distances in order to attempt an interview. The other drawback is that interesting but small sub-groups of the population may not be selected in sufficient numbers to carry out meaningful analysis. It is therefore not used very often.

Systematic sampling is a slight variation on SRS, and shares many of the same strengths and weaknesses. Researchers choose a random starting point on the sampling frame, and then select respondents at fixed intervals. So, for example, if there are 100,000 units listed on the sampling frame, and the aim of the survey is to interview 1,000 people, then the researcher will start at a random position, for example at number 876, and then interview every hundredth person. By and large, this produces the same results as a simple **random sample**. The only difference that may emerge is if there is periodicity in the sampling frame, so that a certain type of person is listed at regular intervals within the sampling frame. For example, if the sampling frame is a database of married couples, and the data is arranged so that women are followed by their husbands, then every second person in the sampling frame will be a man. Accordingly, the 876th person on the list will be a man, as will every hundredth person that follows this starting point. The sample that is drawn will therefore be entirely made up of men.

Probability sample designs can be made better by introducing special features to ensure that the sample adequately represents specific sub-groups of the population that may be of particular interest. Stratified sampling is an example of this. It involves setting different selection probabilities for different groups (or *strata*) in the population. It is frequently used to ensure that adequate numbers of people are selected from different regions of a country. For example, a survey of British adults based on simple random selection might by chance lead to a sample containing very few people from Wales or Scotland, and lots of people from England. To ensure that the sample represents the distribution of the population across these three regions, the sampling frame can be stratified (divided up) into separate lists for each country. Stratifying samples obviously requires that the sampling frame contains information regarding the stratifying variable (e.g. country). The sample can then be drawn from each list. The proportion of the sample drawn from each stratum can either be set to ensure that the sample is representative of Britain as whole or it can vary by strata so that it over-samples from those regions where fewer people live. This allows for the possibility of having a bigger sample size from Wales or Scotland than would otherwise be achieved, which in turn can facilitate the statistical analysis of people living in these regions. These unequal selection probabilities can then

be corrected by **weighting the data** (see Box 11.2) to ensure that the overall sample is representative of Britain as a whole.

Although this method helps to ensure the representativeness of the sample, it does little to ease the costs of carrying out the survey that can make surveys based on simple random selection prohibitive. A technique to try and make this cost more manageable is cluster sampling. With cluster sampling, the units are not sampled individually but in groups (*clusters*). For example, a survey may sample households and then interview every adult within each selected household. Although this saves on transport costs, it increases sampling error, since people living in the same household might share certain characteristics, such as social class, which may not be representative of people living in other households. Multi-stage cluster sampling is an extension of this approach and is often used for large national surveys. This involves drawing several samples. At the first stage, the population is divided up into geographic regions, such as electoral constituencies, which form the clusters. These clusters are then selected using SRS. At the second stage, each of the selected clusters is then divided up into even smaller geographic units, such as electoral wards, which are again selected using SRS. At the third stage, addresses from within each electoral ward are selected, and then within each household a respondent is selected to interview.

BOX 11.2 Weighting the Data

People often talk about weighting the data, and about weighted and non-weighted data, though it is not always clear what this refers to. Broadly speaking, there are two main ways in which this can be done. The most straightforward form of weighting is to correct for unequal selection probabilities. So, for example, if we deliberately over-sample people living in Scotland, Wales, and Northern Ireland (so that we have a sufficiently large sample size to analyse people living in these countries separately), our combined sample will not be representative of the UK population as a whole. To correct this, we apply weights to the data, which adjusts our sample and statistically reduces the number of people living in Wales in our sample by the same degree to which we over-sampled them at the design stage. This brings the proportion of people living in Wales within our sample back into line with the proportion of people living in Wales within the UK. This is a relatively straightforward procedure, since the selection probability is known for every selected case.

The second way in which weights are often applied is to correct for unit non-response. This is more problematic, since it involves estimation. For example, suppose we find that response rates to our survey are higher among older people than younger people. Our achieved sample will therefore be skewed towards older people, and will not be representative of the population as a whole in terms of age. This can introduce bias if young people and old people differ in their political attitudes or behaviour. To compensate for this under-representation of young people, an adjustment is made to our sample which statistically increases the proportion of young people within our sample, so that it matches the proportion of young people in the population as a whole.

The problem with this strategy is that it assumes that non-response is random, and that young people who do not answer the survey are essentially the same as young people who do answer the survey. However, this assumption is not always valid. Young people who do not respond to the survey may be quite different in important respects. They may be less interested in politics, they may be less inclined to give up their time for nothing, or they may simply be occupied with work and going out. In these instances, applying non-response weights will do little to reduce any bias that has crept into the survey.

The advantage of this approach is that typically one investigator can carry out all the interviews within each cluster, since all the selected respondents will be within a relatively confined geographical area.

Non-probability sampling

Everything else being equal, for population surveys, probability samples are always preferred to non-probability samples. Non-probability samples are not only more likely to introduce error into the survey, but they are likely to do so in ways which are difficult to predict. This makes it much more difficult to have confidence in the results. The most common form of **non-probability sampling** is the quota sample.

Quota sampling requires investigators to interview a certain number (quota) of respondents, where quotas are set according to personal characteristics, such as age group, sex, and income group. These quotas are organized to try and get a spread of different types of people, which can then be weighted to ensure that the sample is representative of the population in these respects. Quota samples tend not to make explicit use of sampling frames and so are susceptible to coverage error. Quota sampling is non-random because interviewers can select any subjects they want that fit the selection criteria. This is usually done using some (unstated) form of purposive or convenience sampling, in which investigators approach those people who look most cooperative. Although quota sampling is quite common, and generally gives reasonable results, it is easy to introduce biases in the selection stage, and almost impossible to know whether the resulting sample is a representative one.

Other forms of non-probability sampling are even more unreliable. **Purposive sampling** involves investigators using their own 'expert' judgement to select respondents whom they consider to be typical or representative of the population of interest. It suffers from many of the same problems as quota samples, since it is very easy to introduce bias into the selection. Similarly, **snowball sampling** involves investigators finding respondents who meet some criteria of interest, such as being an animal rights activist or drug user, and then asking the respondents to suggest other similar types of people who they may know for the researcher to contact for interview. In this way, the researcher is able to build up a sample of respondents. However, the problem with this technique is that the cases that are selected are not independent of each other, and are often not independent of the initial starting point from which the snowball was generated. For example, using this sampling strategy, Becker (1963) attempted to interview drug users through his contacts in the music business. Unsurprisingly, the resulting sample was heavily biased towards musicians, who made up half of his sample. It was not therefore very representative of the population of interest. Snowball samples tend to be very unreliable, and are not a sound basis for making wider inferences. However, they are sometimes used in qualitative research, particularly when the population of interest is hard to reach, and does not have a reliable sampling frame (see Chapter 12).

However, perhaps the most unreliable of all sampling strategies is the **volunteer sample**. This is the sort of survey that is frequently carried out by morning television shows and internet sites, where viewers or readers are encouraged to express their opinions on various

topical issues of interest. For example, consider what inferences you can draw from the following example of a Sky News viewers' poll:

Should Britain join the space race?

No—It's cosmic showboating	52 per cent
Okay, but let's spend sensibly	30 per cent
Yes—I'd be over the Moon!	18 per cent

Putting aside issues to do with question wording, it tells us very little. We would be unwise to infer that the British public is against joining the space race. We would even be unwise to infer that Sky viewers are against joining the space race. All we learn from this type of exercise are the opinions of those readers or viewers who could be bothered to send in their response. For these reasons, volunteer samples (even the more serious ones) tell us essentially nothing about the average attitudes of the general population.

Sampling error

Even though probability samples tend to be more representative of the population than non-probability samples, they are still affected by different types of sampling error. Sampling error can occur for a variety of reasons. Two main factors that are important are to do with the response rate that the survey achieves and the sample size of the survey.

Ideally, everyone that is sampled will complete the survey. However, this rarely happens, and non-response can occur for a variety of reasons. Some people will refuse to participate, or will be unable to participate because they are too sick or because they are away on business. Non-response may therefore reduce sample size and introduce bias. Those people who do not answer the survey may be somewhat different in some key respects from those people who do answer surveys. If someone is not interested in politics, they might not want to spend an hour answering questions about the most recent election. Similarly, if someone is too sick to be interviewed, they might be too sick to do lots of other things as well, such as go out to vote.

An example of this relates to the measurement of turnout. It is not uncommon to find that the proportion of respondents who report voting in surveys far exceeds the official estimates of actual voter turnout. For example, in the 2017 BES, self-reported turnout was measured at 79 per cent, well above the official turnout figure of 69 per cent. This substantial discrepancy between survey estimates of self-reported turnout and official estimates of turnout is not just limited to Britain, it is common to most surveys carried out around the world (and has been a particular source of concern in, for example, the United States). A study by Swaddle and Heath (1989) examining the sources of this discrepancy found that a considerable portion of it was due to problems of response bias (where turnout among non-respondents is substantially less than turnout among survey respondents).

Non-response may introduce bias if those individuals who do not agree to be interviewed for the survey (or cannot be interviewed because they are away) are less likely to vote than those individuals who agree to be interviewed. Swaddle and Heath found that, according to the official turnout register, turnout among the survey respondents (83 per cent) was

substantially higher than among the non-respondents (75 per cent). A substantial portion of the apparent over-report in turnout by surveys is thus attributable to issues of response and coverage. One obvious way in which survey measurement can be improved is therefore to ensure a high response rate. However, this is not easy to achieve, and in recent years response rates for face-to-face surveys have fallen to around 60 per cent. Although researchers can try to correct for **non-response bias** by weighting the data, this is not a perfect remedy and does not get rid of the problem.

Survey organizations thus devote considerable time and energy trying to maximize response. Various strategies are employed, such as sending an advanced letter to the respondents, visiting the respondent in person, with repeat visits on different days at different times if the respondent is unavailable. One strategy that is becoming more common is the use of inducements, such as a cash incentive or gift. It is unclear how much difference these inducements actually make though. Research on different types of inducement show that entering the respondent into a prize draw has no impact on response, whereas an unconditional payment (whether they complete the survey or not) is most effective. However, this can obviously increase the costs of the project substantially.

How you ask: interview mode

The last stage of survey design that we consider is how the survey is administered. This is an important part of data collection. No matter how well a survey is designed, if it is poorly executed, then it will produce unreliable results. Over the past twenty-five years or so there has been a rapid expansion in the ways—or modes—that surveys can be carried out. The most common ways are face-to-face, by telephone, by post, and, more recently, by internet. The way in which the survey is carried out can influence both how people answer the questions, relating to issues of measurement error, and who answers the questions, relating to issues of sampling error (see Groves et al. 2004 for an extended discussion). We briefly describe the main differences between each of these modes of data collection, and then discuss some of the issues they raise. Whatever mode is used, it is important that respondents are informed about the survey and that the survey is conducted in an ethical way (see Box 11.3).

BOX 11.3 Ethical Issues in Survey Research

The key ethical issues in survey research relate to informed consent, voluntary participation, and privacy. Before the survey starts, respondents should be told what the survey is about (in general terms), who is conducting the survey, and for what purpose (informed consent). They should also be told that they do not have to answer any questions and can stop the survey at any point (voluntary participation). Lastly, they should be assured that any answers they provide will be treated in the strictest confidence and all the results will be anonymous (privacy). Protecting the privacy of respondents doesn't just mean that all the results should be anonymized. It also means that respondents should not be identifiable from the answers they have provided. In particular this means that details about where the respondent lives should not be disaggregated to a level where it might be possible to identify them, such as by including their postcode in the final results. For more details, see Social Research Association (SRA) *Ethical Guidelines* (2003).

Face-to-face interviews remain the gold standard in survey research, and most of the big academic and government surveys are carried out in this way. However, they are very expensive. Respondents are interviewed by trained investigators who administer the questionnaire personally, usually at the respondent's home. Investigators read out the questions to the respondent, show prompt cards with the responses (for closed questions) and record the verbal answers that the respondent provides. Traditionally, these interviews were carried out with pen and paper, but now they tend to be carried out using laptops (computer-assisted personal interviewing—CAPI), which can be used to play video clips and speeches. Face-to-face surveys can clearly establish who is answering the questions, and whether they are doing so freely and comfortably without intervention from other people. Telephone interviews are also administered personally by trained investigators, though obviously in a more remote way since they take place over the phone. They do not provide response cards and so lack visual cues. By contrast, both *postal surveys* and *internet surveys* are self-administered by the respondent, who must fill out written versions of the questionnaire by themselves. Internet surveys in particular are becoming more popular, and are frequently used in academic research. They have the advantage of being relatively cheap to administer, and there is now a whole host of websites on the internet which can be used to host the survey at relatively little cost, which can make it seem an attractive option for researchers with a limited budget.

Besides budget, there are a number of other important factors that need to be considered. The way in which a questionnaire is administered may influence the way in which respondents answer the questions in a number of different ways. First, face-to-face surveys (and to a lesser extent telephone surveys) can help to alleviate **comprehension problems**. The interviewer can make sure that the respondent has understood the question properly, and can repeat a question if necessary. Since personally administered interviews are carried out verbally, they also place less of a burden on the literacy skills of the respondent. By contrast, people with reading and writing difficulties may feel less confident and be less able to complete self-administered surveys.

However, there is another side to this. The presence of an investigator can in some cases inhibit the responses that are given. These are known as **interviewer effects**. To illustrate, Schuman and Converse (1971) carried out a study on whether the race of the interviewer influenced respondents' answers towards questions on race relations. They found consistent interviewer effects, whereby black respondents were much more likely to say that they trusted white people if they were interviewed by a white investigator than if they were interviewed by a black investigator. The more anonymous setting of internet and postal surveys may therefore help to alleviate these problems, and in the process may help to reduce social desirability bias which can affect certain sensitive questions. For these reasons, for sensitive topics, face-to-face surveys using CAPI often pass the laptop to the respondent and ask them to answer the questions on the screen by themselves. With respect to measurement issues then, the mode of interview that is preferred may vary, and depend in part upon the type of questions that are being asked in the survey, and whether they are complex on the one hand or sensitive on the other.

The way in which a questionnaire is administered can also influence the representativeness of the sample. In particular, problems to do with coverage error and non-response may vary by interview mode (see Groves et al. 2004). The best way to ensure that a sample

is representative of the population is to draw a random sample, in which all members of the population have a known non-zero probability of being selected. To achieve this it is necessary to define the population and to select an appropriate sampling frame from which to draw the sample. However, the coverage of common target populations varies by different methods of data collection, and even where coverage is high, it is not always possible to find a reliable sampling frame. For face-to-face surveys and postal surveys, the most comprehensive national sampling frame from which to draw a sample is the Post Office Address file (or equivalent). This type of sampling frame tends to cover pretty much all the addresses in a country and so does not introduce much coverage error. However, obtaining a sampling frame with good national coverage is much more problematic for telephone interviews and internet interviews, and this can introduce greater coverage error.

For telephone surveys, even though the proportion of households with a telephone tends to be quite high in developed countries (typically over 90 per cent), it is nonetheless difficult to obtain a comprehensive sampling frame. For example, the phone book (Yellow Pages or equivalent) is not very representative of the population because of the large number of people who are ex-directory (not listed). For example, in the UK about 25 per cent of the population who have a telephone are not listed in the phone book. Moreover, the rise of mobile phones mean that fewer people today have a fixed landline than in the past, and those that do have a landline tend to be somewhat older. The phone book is therefore not a particularly good sampling frame, and researchers who do telephone surveys tend to use **random digit dialling** instead. This involves dialling numbers at random. Although this reduces coverage error (though the coverage of mobile phone numbers is patchy) it involves making a lot of wasted calls to numbers that aren't in use, and does not distinguish between business and residential addresses. This can make it difficult to calculate response rates (as we will discuss).

For internet surveys, the past few years have seen a rapid increase in the proportion of British households with internet access, and according to the Office of National Statistics, about 70 per cent of British households now have internet access. But this is still a long way away from full coverage, and any survey carried out through the internet will leave out a substantial proportion of the population. Moreover, the so-called 'digital divide' between those who are online and those who are not means that there are also likely to be important demographic differences between the two groups, which may introduce further bias (see Robinson, Neustadtl, and Kestnbaum 2002). Moreover, for internet surveys, no good sampling frame has been developed for sampling the internet population (see Couper 2000 for a review). Survey companies therefore often rely upon drawing samples from a database of volunteers, which can then be weighted to make the sample representative of the population with respect to key demographic variables. This sampling strategy mixes elements of the volunteer sample and the quota sample, discussed previously.

Despite these problems, internet and telephone surveys are frequently used by commercial polling agencies, and are even widely used in academic research. In particular, the low cost of internet surveys makes them a very attractive option when compared to the far more expensive face-to-face surveys. Although some research suggests that there is little observable difference between analyses using in-person data and internet data (see Sanders et al. 2007) the methodology employed by companies carrying out internet and telephone surveys has come under severe scrutiny following the botched attempt to predict the 2015

UK General Election. The pre-election polls—which were all based on telephone and internet surveys—significantly underestimated the size of the Conservative lead over Labour at the 2015 General Election, with nearly all the final polls predicting a dead heat. The failure to predict the Conservative majority led many commentators to question the value and robustness of the polling methodology, with some even calling for polls to be banned in the final weeks preceding an election. In response to these events, the British Polling Council and the Market Research Society launched an independent inquiry into the causes of the discrepancy between the final polls and the election result (see Box 11.4).

Indeed, in recent years there has been a growing sense that polls are not as accurate as they once were. A number of high profile cases, such as Brexit and the election of Donald Trump, where the polls seemed to suggest the wrong winner have reinforced this belief. But is it true? In order to answer this question Will Jennings and Chris Wlezien (2018) collected information on 26,000 published polls in 338 elections from forty-five countries since 1942 and compared the polling estimates to the official election results. The 'glass-half-full' interpretation of their findings are that polls are not getting any worse over time, and are roughly as good as they have ever been. The 'glass-half-empty' interpretation though is that despite nearly eighty years of accumulated experience, and massive technical and statistical developments, there is not much evidence that polling has got any better.

One of the main challenges for all types of survey research is how to obtain a high **response rate**, since non-response can introduce error if people who do not complete the survey differ in key respects from those who do (Prosser and Mellon 2018). It is therefore important to try and achieve a high response rate, and also, to be aware of who does not respond to the survey and why. Traditionally, response rates for population surveys have tended to be higher for face-to-face surveys than for telephone surveys and mail and

BOX 11.4 Why Did the Opinion Polls Get It Wrong?

In the run up to the 2015 British elections, all the major opinion polls were predicting a hung parliament. In the last week of the election campaign, the polls had the Conservative Party and the Labour Party neck and neck, with some polls indicating a slight lead for Labour. When the results came in they were a huge shock for everyone. The Conservatives were not only the largest party, but managed to secure an overall majority, winning nearly eight percentage points of the vote more than Labour. Why then did the polls get it so wrong? Why did they over-estimate the Labour share of the vote and under-estimate the Conservative share of the vote? The answer is almost certainly to do with how the polling companies selected their samples. All of the polling companies used either telephone or internet surveys, which relied on a combination of quota and what effectively amounted to volunteer samples. As a consequence the samples were not representative of the British electorate, even after they had been weighted. In particular the preliminary findings of UK Polling Inquiry found that the samples were 'very politically engaged', meaning that the people who took part in the internet and telephone surveys were much more interested and active in politics than the general population. This created a problem as political engagement is also associated with voting behaviour. By contrast, the more rigorous face-to-face surveys based on random probability sampling carried out by the BES and British Social Attitudes shortly after the election contained much more representative samples, and much more accurate predictions of vote preferences. For more details, see Sturgis et al (2018), the BPC/MRS Polling Inquiry website and the BES website.

internet surveys (de Vaus 2002; Groves et al. 2004). However, there is also substantial variation across modes of interview depending upon how much effort is made to re-contact respondents, and according to Groves et al. (2004: 154), it is not clear whether there are any inherent differences between the modes that directly affect response rates.

Mode effects are more apparent when it comes to identifying the reasons for non-response, and for distinguishing between different types of non-response. For example, with face-to-face interviews, it is possible to distinguish between eligible units and ineligible units. An ineligible unit might be a business address or a derelict property where no one lives. In general, we are only concerned about the non-response of eligible units, so this information is important. However, in mail surveys it can be hard to distinguish between the two. If there is no response, it may be because the potential respondent did not complete the survey (eligible unit), or it may be because the survey was sent to a derelict property where no one lives (ineligible unit). Similarly, with telephone surveys, if the telephone rings but no one answers, it may be because the potential respondent is away or busy (eligible unit) or it may be because the telephone number is not in action (ineligible unit). Lastly, although emails sent to bad addresses are frequently bounced back, it is not always possible to tell whether it has reached its intended target.

In summary, although there are advantages and disadvantages in using each of the different interview modes, these are not fixed and vary according to the purpose of the survey and the target population. Moreover, modes of interview can be combined. For example, it is common practice to supplement face-to-face surveys with a mail back or internet follow-up, which can be a cheap and reliable way of gathering more data once the initial sample has been drawn.

Conclusions

As the examples in this chapter have shown, there are a wide range of difficulties that face researchers trying to undertake survey research. These difficulties come in many forms and can introduce error at various stages of the research process. But despite these difficulties, and perhaps because we are aware of them, surveys remain a valuable source of information. Some difficulties are easier to manage than others. By using random probability sampling, we can have confidence that our surveys are more representative of the population at large than surveys conducted by any of the other alternatives. And this representativeness can give us confidence that inferences we make from our sample are generalizable to the wider population. But ensuring representativeness is only part of the challenge.

Questionnaire design is not an exact science and the researcher is confronted with many measurement problems. Some of these problems can be at least partially remedied by careful question wording, but other problems persist and are harder to get rid of. Questionnaires are a very delicate instrument. Slight changes in how questions are worded or phrased can influence how people answer. This means that we should be very careful how we interpret the results from different survey questions. For example, if we are interested in the balance of public opinion for or against accepting refugees from Syria, we have to bear in mind that our survey estimate is influenced in part by how we have asked the question. A similar

question framed in a slightly different way may produce a different estimate. If we want to know whether public opinion has changed, we therefore have to be careful to only compare survey questions that have been asked in exactly the same way. Even small changes in question wording could make the observed differences an artefact of measurement differences, rather than an indication of any real difference of opinion. For these reasons, when we analyse survey data, we tend to focus on whether responses to a particular question vary over time; or whether they vary by different sections of society.

Questions

1. What are the principles of good questionnaire design? Are these principles generally followed in survey research?
2. If the way in which we ask questions affects the answers that we get, can surveys ever be trusted?
3. What is measurement error? Can it be avoided? If not, what should we do?
4. What is the difference between validity and reliability?
5. How do we know if a survey question is valid?
6. What is the difference between measurement error and bias? Which should we be more concerned about? Why?
7. What are the strengths and weaknesses of different sampling techniques?
8. Are non-probability samples really so bad? Why?
9. What are the strengths of survey research? What are the weaknesses?
10. What are the different types of error that affect survey research? What can be done to try and reduce them?
11. As long as we are aware of the error we make, it doesn't matter if we make errors. Do you agree?

Guide to Further Reading

de Vaus, D. (2002), *Surveys in Social Research* (London: Routledge).
A detailed and accessible overview of survey design and administration. Contains many easy-to-read and helpful chapters on how to carry out a survey and how to analyse the results.

Groves, R., F. Fowler, M. Couper, J. Lepkowski, E. Stinger, and R. Tourangeau (2004), *Survey Methodology* (New Jersey: John Wiley and Sons).
A sophisticated examination of survey error and how to minimize it.

Moser, Sir C. and G. Kalton (1971), *Survey Methods in Social Investigation* (London: Heinemann Educational Books).
A classic book on survey design. Provides a thorough examination of survey research, and offers advice on how to plan, design, and carry out a survey effectively.

Schuman, H. and S. Presser (1996), *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context* (London: Sage Publications).
A fascinating study reporting the findings from a number of survey experiments designed to investigate the impact of questionnaire design on the responses that are given.

References

Atkeson, L. (1999), '"Sure, I Voted for the Winner!" Overreport of the Primary Vote for the Party Nominee in the National Election Studies', *Political Behavior* 21(3): 197–215.

Becker, H. S. (1963), *Outsiders: Studies in the Sociology of Deviance* (New York: Free Press).

Belli, R., M. W. Traugott, M. Young, and K. A. McGonagle (1999), 'Reducing Vote Overreporting in Surveys', *Public Opinion Quarterly*, 63: 90–108.

Belson, W. (1981), *The Design and Understanding of Survey Questions* (London: Gower).

Converse, P. E. (1964), 'The Nature of Belief Systems in Mass Publics', in D. Apter (ed.), *Ideology and Discontent* (New York: Free Press of Glencoe), 206–61.

Couper, M. (2000), 'Web Surveys: A Review of Issues and Approaches', *Public Opinion Quarterly* 64: 464–94.

de Vaus, D. (2002), *Surveys in Social Research* (London: Routledge).

Gowers, E. A. (1954), *The Complete Plain Words* (London: Penguin).

Groves, R., F. Fowler, M. Couper, J. Lepkowski, E. Stinger, and R. Tourangeau (2004), *Survey Methodology* (New Jersey: John Wiley and Sons).

Heath, O. and R. Johns (2010), 'Measuring Political Behaviour and Attitudes', in M. Bulmer (ed.), *Social Measurement through Social Surveys: An Applied Approach* (Aldershot: Ashgate), 47–68.

Heath, O., P. Smith, N. Gilby, and F. Hoolahan (2015), 'Partisan Priming and Subjective Evaluations of the Economy: Evidence from a Survey Experiment', *Journal of Elections, Public Opinion and Parties* 24(4): 530–42.

Jennings, W. and C. Wlezien (2018), 'Election Polling Errors across Time and Space', *Nature Human Behaviour* 2(4).

Moser, Sir C. and G. Kalton (1971), *Survey Methods in Social Investigation* (London: Heinemann Educational Books).

Prosser, C. and J. Mellon (2018), 'The Twilight of the Polls? A Review of Trends in Polling Accuracy and the Causes of Polling Misses', *Government and Opposition* 53(4): 757–90.

Robinson, J. P., A. Neustadtl, and M. Kestnbaum (2002), 'The Online "Diversity Divide": Public Opinion Differences among Internet Users and Nonusers', *IT & Society* 1: 284–302.

Sanders, D., H. Clarke, M. Stewart, and P. Whiteley (2007), 'Does Mode Matter for Modeling Political Choice? Evidence from the 2005 British Election Study', *Political Analysis* 15: 257–85.

Schuman, H. and J. Converse (1971), 'The Effects of Black and White Interviewers on Black Responses in 1968', *Public Opinion Quarterly* 35(1): 44–68.

Schuman, H. and S. Presser (1996), *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context* (London: Sage Publications).

Schuman, H., J. Ludwig, and J. Krosnick (1986), 'The Perceived Threat of Nuclear War, Open Questions, and Salience', *Public Opinion Quarterly* 50: 519–36.

Social Research Association (2003), *Ethical Guidelines*. Available at: http://the-sra.org.uk/wp-content/uploads/ethics03.pdf

Sturgis, P., J. Kuha, N. Baker, M. Callegaro, S. Fisher, J. Green, W. Jennings, B. E. Lauderdale, and P. Smith (2018), 'An Assessment of the Causes of the Errors in the 2015 UK General Election Opinion Polls', *J. R. Stat. Soc. A* 181: 757–81.

Sudman, S. and N. M. Bradburn (1982), *Asking Questions* (San Francisco, CA: Jossey-Bass).

Swaddle, K. and A. Heath (1989), 'Official and Reported Turnout in the British General Election of 1987', *British Journal of Political Science* 19(4): 537–51.

Tourangeau, R., L. J. Rips, and K. A. Rasinski (2000), *The Psychology of Survey Response* (Cambridge: Cambridge University Press).

Wright, G. C. (1993), 'Errors in Measuring Vote Choice in the National Election Studies, 1952–88', *American Journal of Political Science* 37: 291–316.

Zaller, J. (1992), *The Nature and Origins of Mass Opinion* (Cambridge: Cambridge University Press).

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

Interviewing and Focus Groups

Chapter Summary

One of the ways in which researchers collect information about the social world is to ask people questions. This can be done by conducting structured and semi-structured interviews, face-to-face interviews, telephone or online interviews, or organizing focus group sessions. This chapter discusses different types and forms of interviewing, how interviews and focus groups (a form of interview) should be carried out, and coding and analysing interview data. The questions addressed in this chapter include the following:

- What are the different types and forms of interview?
- How can interview data be used to confirm or disconfirm a hypothesis or argument?
- How do you plan and conduct an interview?
- What are the different types of questions that might be included in an interview?
- How does the type and wording of questions, and the order in which you ask them, affect the responses you get?
- What interviewing skills will ensure a more successful outcome to an interview?

Introduction

Interviews are a prominent method of data collection in political research. *Interviews* as a method of collecting data in research can be distinguished from *surveys*, and also from interviews that are conducted for purposes other than research (such as job interviews or media interviews). There are similarities between interviews and surveys (this is discussed in Chapter 11). Both use questions to find out answers. But these similarities relate mostly to structured interviews. This chapter focuses on semi-structured interviews (including focus groups). In these types of interviews often the researcher wants to probe, which is rather like asking leading questions, to get the interviewee to open up and discuss something of relevance to the research question. This kind of technique is generally not used in survey research. The role of the interviewer is therefore somewhat different, and in an interview the interviewer is more active. More generally, we can say that the aims of interviewing and surveys can also be somewhat different

(although they are frequently complementary). The aim of *surveys* is to produce standardized data which can be used to make generalizations about what a given population of people think about a particular issue. By contrast, *interviewing* is more concerned with obtaining detailed, often specialized information from a single individual or small number of individuals. The aim is not necessarily to make generalizations, but to gain valid knowledge and understanding about what the person in question thinks. Although both interviews and surveys take the issues of reliability and validity seriously, it may be argued that surveys tend to prioritize reliability (which allows researchers to make generalizations), whereas interviews prioritize validity (which allows them to gain a depth of knowledge).

In this chapter we will discuss the different types and forms of interviews, what considerations enter into designing a schedule of questions, how to conduct an interview and organize a focus group, and how to analyse interview data.

Interviews

To decide whether interviews will enable you to investigate your hypothesis you must ask yourself—and answer—a number of questions. How will interviews enable you to investigate/demonstrate your hypothesis? What, specifically, do you need to find out through interviewing to enable you to investigate the relationships stated by your hypothesis? Is conducting interviews the best way to obtain this information? What kind of interview, and what sort of questions, will enable you to obtain this information? Of whom will you ask questions and what techniques will you use? Understanding the various types and forms of interview will enable you to answer these questions.

Types of interview

Different *types of interview* provide different ways of collecting interview data. Interview data can be collected by conducting individual face-to-face, telephone, or online interviews, or by organizing focus group sessions. These types of interview vary with respect to the degree of personal contact they entail between researchers and subjects. However, they all involve a situation in which a researcher asks respondents questions and then records and analyses their answers. For all types of interviewing, the primary purpose is to produce data that will help answer research questions.

Individual face-to-face interviewing

This type of interview is the best data-collection type for open-ended questions and in-depth exploration of opinions. The interviewer works directly with the respondent and has the opportunity to probe or ask follow-up questions. By being able to ask questions of subjects personally, the interviewer can probe unclear responses, resolve difficulties that lead to non-responses, and obtain useful information from body language and vocal cues. In sum, this type of interviewing can support longer and more detailed questions, is adaptable, and is a rich and dense source of data.

Telephone interviews

This type of interview enables a researcher to gather information rapidly. Most major public opinion polls that are reported are based on telephone interviews (see Chapter 11). Using telephone interviews gives researchers wide geographical access: people from all over the globe can be interviewed if they have access to a telephone or computer, as well as populations that are difficult to reach or interview by other means (e.g. people living in remote areas). Like face-to-face interviews, they allow for some personal contact between the interviewer and the respondent, and they allow the interviewer to ask follow-up questions. Telephone interviewing supports longer questionnaires than mailed questionnaires, though not as well as face-to-face interviewing.

Online interviews

Online interviews are increasingly used as a data-collection method by social scientists. These can be a useful technique for data collection on sensitive items, and they eliminate interviewer bias. But online interviews are a far less personal type of interview than other types; they are not an appropriate tool for exploratory studies, which require intensive interaction with interviewees in order to gain better insight into an issue. They can be useful in small-scale studies on sensitive topics and in combination with other tools. This is a quick and easy way to reach geographically dispersed interviewees.

Focus groups

Focus groups generate different types of data from other types of interviewing. In this type of interview, a group of people is selected because they are related to some phenomenon of interest. The researcher meets with the group and facilitates an organized discussion related to something the participants have experience of, or beliefs about. The interaction among focus group participants may reveal more than they would in an individual interview, and bring out insights and understandings in ways which simple questionnaire items (used in face-to-face or telephone interviews or sent through the mail) may not be able to tap into, such as emotional and unconscious motivations. Using focus groups, researchers can learn how things are discussed in a particular culture, test hypotheses about beliefs, bring together people who wouldn't normally mix, test questions for future interviews, or test policy ideas to see how citizens react to them.

Advantages and disadvantages of different types of interview

Let's compare the advantages and disadvantages of different types and forms of interviewing for specific purposes. Individual face-to-face interviews or focus groups provide more flexibility and can enable researchers to probe or ask follow-up questions. However, they always carry the possibility of interviewer bias, as well as respondent bias on sensitive items. They require skill to undertake well; they can also be very expensive, take a lot of time, and will usually be less feasible to do if the population you wish to interview is widely dispersed

geographically. The data from both types are difficult to analyse—particularly in the case of focus groups, where the analysis may be concerned both with the content of what was said and the patterns of interaction.

Telephone interviews enable the researcher to interview people who are not easy to access, but the interviewer does not see the interviewee, so body language and voice and intonation cannot be used as a source of extra information. Moreover, the interviewer has no control over the setting and ambience and cannot standardize the situation in which the interview takes place. And telephone interviews have to be relatively short or energy levels will flag. Telephone interviews are not ideal if you want to ask lengthy questions, give the respondent detailed background for a question, or ask the respondent to select from a long list of possible options or answers.

Online interviews have become increasingly popular. In cases where it is not possible to schedule a time to meet face-to-face or to speak on the phone, it might be the *only* way to conduct an interview. Online interviews present opportunities to interview individuals based anywhere in the world. And they provide more time for interviewer and interviewee to consider their questions and answers, allowing for increased flexibility and deeper reflection. Unlike more personal forms of interviewing (individual face-to-face, focus group, or telephone interviewing), they allow the interviewer to ask questions that may require the respondent to consult records in order to supply answers. However, they are a poor choice for complex questions, and the researcher cannot probe or collect additional data. As with telephone interviewing, the researcher has no control over whether or how a respondent is interacting with others while completing the questionnaire.

Forms of interview

You will not be able to select among these different *types* of interviewing—to fully weigh their advantages and disadvantages for different purposes—unless you are clear about how, within these types of interview, the *form* of interviewing can vary. What we want to clarify is: what are the different formats that interviews take, including the types of questions asked and the manner in which they are asked? Which of these forms or formats are supported by the different types of interviews that we just discussed?

There are three basic *forms of interview* that vary with respect to the degree to which the questions and responses are standardized across interview subjects. These are *structured*, *unstructured*, and *semi-structured* forms of interview. They differ with respect to the form of question employed (closed or open), how they are worded (shorter and simpler, or longer and more complex), and whether the same questions are asked in the same way and in the same sequence for each interview subject (standardized or variable). **Structured interviews** consist of a standardized set of closed and shorter or simpler questions that are asked in a standardized manner and sequence. This form of interview is primarily used in surveys, and is discussed in detail in Chapter 11. Of the types of interviewing we discussed previously, mailed questionnaires almost invariably use this form of interviewing. **Unstructured interviews** use open, and perhaps lengthier and more complex questions, which might vary in the way and order in which they are asked. Focus groups, which involve interviewing people together in flexible and exploratory group discussion formats, are an example of a type of interview that uses this form of interview.

Perhaps the most common type of interview used in political research is the **semi-structured interview**: a form of interview that combines elements of both structured and unstructured interviews.

Unstructured interviews

Unstructured interviews often use open questions, which allow for longer and perhaps more complex questions, as well as more in-depth probing and flexibility, than structured interviews. They are more like ordinary conversations: there is no set interview structure, and interviewees answer in their own words. The interviewer initiates the conversation, presents each topic by means of specific questions, and decides when the conversation on a topic has satisfied the research objectives. These elements of the interview process are likely to vary with each individual interview, based on the responses of the interviewee.

Semi-structured interviews

Semi-structured interviews generally involve a small number of interviews in which the interviewer uses a combination of structured questions (to obtain factual information) and unstructured questions (to probe deeper into people's experiences). The interviewer prepares a schedule of questions, as well as supplementary questions to explore aspects of the response to each question asked. But the questions may be more general, and the interviewer may choose to vary the sequence in which they are asked, and to ask follow-up questions. Unstructured or semi-structured interviews can give greater insight into the meanings of a subject's experiences and hence provide more valid data. However, it is not as easy to compare responses of different groups or of the same group over time since each interviewee will have been asked slightly different questions to find out what they think. The data are not standardized and are thus hard to generalize from and, as the results cannot be quantified and retested, less reliable.

Box 12.1 shows how these types and forms of interview commonly combine in political research.

Which type and form of interview to use?

When selecting the type of data-collection technique and form of interview to use, there are a number of substantive and practical issues to consider.

Generally, structured interviews are better for making comparisons and less structured interviews may be more appropriate for early exploratory phases of research. Ultimately, it will be the nature of what you want to ask respondents that determines the type and form of interview you select. Are you going to need to get lots of detail in the responses (open questions); or can you anticipate the most frequent or important types of responses and develop reasonable closed-ended questions? Will the questions be complex and require multiple parts or sub-questions? Can you construct in advance a set of questions and the sequence in which they will be asked? Or will you need to proceed in a more exploratory manner, perhaps entailing lots of follow-up questions that you can't easily anticipate?

BOX 12.1 Types and Forms of Interviews

	Forms of interview (types of questions, wording, and sequencing)		
	Structured (closed, shorter, standardized)	Semi-structured (combines structured/ unstructured forms)	Unstructured (open, longer/ more complex, varied)
Types of interview (data-collection techniques)			
Individual face-to-face	*x*	*x*	*x*
Focus groups			*x*
Telephone	*x*	*x*	*x*
Online	*x*	*x*	*x*
Mailed questionnaire	*x*		

A key consideration is the likelihood of bias and of threats to the validity and reliability of your findings. A limitation with all types and forms of interviews is that researchers are dependent on what people tell them. Some people may have problems in recalling information accurately. But all people come to an interview with biases and prejudices; and people are generally prone to something known as the 'interview effect': the tendency for interviewees to give more 'socially acceptable' answers or answers they think the interviewer wants. The biases and prejudices of interviewers may also distort the interview process: they may not ask questions that make them uncomfortable, or fail to listen carefully to respondents on topics on which they have strong opinions. People may not tell the truth because they want to 'look good' in the eyes of the interviewer, or if they are asked something that either they don't know how to answer or that would be embarrassing for them to answer. Some types of interview work better in reducing the effects of biases and prejudices. Anonymous questionnaires that contain no identifying information are more likely to produce honest responses than those that identify the respondent.

In selecting the type of data-collection technique, there will be a number of practical considerations, such as cost, facilities, equipment, and time. Cost is often a major determining factor in selecting the type of interview to conduct. Costs can include travel expenses entailed in conducting personal interviews, or postal costs for an extensive mailing. Facilities are another consideration. You will require access to well-equipped phone-surveying facilities for telephone interviews, and a comfortable and accessible room to host focus groups. Equipment needs might include a transcription machine with a pedal-operated start–stop mechanism, or voice transcription software—such as *Dragon Naturally Speaking* speech recognition software from Nuance Communications—to record responses, and equipment or resources to pay for professional transcription.[1] Finally, will you have enough time to carry out all the personal interviews?

An interviewer conducting unstructured or semi-structured interviews will generally prepare an **interview guide**: a set of questions that provide a list of talking points or areas to be covered. We will now discuss this approach to collecting data.

Designing an interview guide for semi-structured or unstructured interviews

A semi-structured interview is a powerful research instrument which can help a researcher understand people's perceptions, feelings, opinions, experiences, understandings, values, beliefs, attitudes, emotions, behaviour, formal and informal roles, and relationships. In semi-structured interviews, the interviewer guides the interview, but permits various aspects of the topic to arise naturally and in any order. To do this, researchers will usually construct an 'interview guide': a basic checklist of areas to be covered in the interview in the form of questions. But, although the interviewer comes prepared with a list of areas to be covered, and may prompt and give direction to the interviewee, the interviewee is treated as an active subject, and not merely a reporter of facts or experiences.

Take a look at this draft of an interview guide on family life in border areas.

DRAFT INTERVIEW GUIDE

Introduction: purpose of research, tape recorder/confidentiality, conduct of interview.

BIOGRAPHY

1. *Tell me about your family background. What was it like when you were a child?*

   Prompt: Level/main source of household income

   Education of family members

2. *Are there holidays or special family events that your family celebrates?*

   Prompt: Could you tell me more about these?

3. *Tell me something about your extended family.*
   Probe: What sort of connections did your family maintain with this larger group of family members while you were growing up?
   How many family members live across the border?

FAMILY PRACTICES

4. *Does your family cross the border to see family members?*
5. *What is the pattern of your cross-border family visiting? How often do you cross the border to see family members?*

   Probe: Places you visit

   Usual length of visit

6. *Do you celebrate holidays and special occasions with members of your extended family who live across the border?*
7. *Do family members keep in touch with members across the border in other ways?*

CROSS-BORDER ACTIVITIES

8. *Do you cross the border to participate in family celebrations?*

   Prompt: Which ones?

BOX 12.2 Example of an Interview Guide on Social Movements

Overarching themes	Questions	Possible probes (follow-up questions)
Personal background	Please can you tell me a little about your own participation in the movement?	• How long have you been a member? • What led you to join? • Have you remained an active member continuously since the time you joined?
	Can you tell me what types of meetings you have attended?	• How many times have you attended? • Which ones do you find you have liked most? • Why was that?
About the movement	If you think about the movement, what would you say its major accomplishments have been?	• Raising awareness of key issues? • Encouraging wider participation in public life? • Raising funds for key activities and services?
	What do you think are the main strengths of the movement?	• Differences between the movement and other similar movements? • What has enabled the movement to develop these strengths?
	What do you think are the main challenges that the movement faces?	• What makes you think so? • Do you think others in the movement would agree?
Perceptions about the other members	In general, why do you think people become active members of the movement?	• What type of person becomes a member? • Have you ever talked with other members about this?
	Do you see members outside of meetings?	• Why not? • What sorts of things do you do or talk about with other members?
Perceptions about recent activities	Do you think the movement's recent campaign was successful in achieving its aims?	• Do you think anything should have been done differently? • How do you think the membership at large felt about the campaign?
	Do you think the movement should make any changes in how it promotes its interests?	• If you had a leadership position in the movement, would you have done anything differently?

9. *Do you cross the border for other purposes?*

   Prompt: Can you tell me briefly about these?

10. *Have there been any marriages in your family between individuals living on opposite sides of the border?*

    Prompt: Was the marriage considered a special marriage or was it treated like any other family marriage?

EXPERIENCE OF CROSS-BORDER VISITS

11. *What is it like travelling across the border? Does it take a long time to pass through the border control station?*
12. *What are your impressions of the places you have visited across the border?*

    Probe: What do you like/dislike about these places?

Another example of how an interview guide on social movements might be constructed is shown in Box 12.2.

In conducting a semi-structured or unstructured interview, the aim of the interviewer is both to ask questions and to listen in a way that encourages the interviewee to talk. In particular, the interviewer tries to adopt a tone and manner that gives interviewees the freedom to open up, reflect, and express themselves. As interviewer, you must make sure not to lead the interview in a way that closes off discussion by the manner in which you ask questions and receive responses. Here are some general 'rules of thumb':

- Avoid questions phrased in a way that suggests or assumes a particular kind of answer. For instance, rather than asking what the respondent likes (or dislikes) about a candidate, an interviewer might ask, first, how the respondent feels about the candidate, and then follow up with a question about what the respondent most likes (or dislikes) about the candidate.
- Don't discourage long answers; instead, verbally and through body language encourage the interviewee.
- If you don't understand something, continue to listen rather than interrupt. You can ask for clarification during a break in conversation. Probe in ways that indicate your interest in what the person is saying and encourage them to explain in more detail.
- Don't communicate, either verbally or through body language, your judgement about what the interviewee has told you. Make sure that your body language does not appear judgemental, or expose any negative judgement. Don't react to provocative statements but accept what the interviewee says.
- Use body language that signals interest (focus on interviewee, eye contact, nodding, smiling).

A semi-structured or unstructured interview will generally follow the same sequence:

- *Introduction*. The aim of the introduction is to explain the research and the interview process, and to establish rapport with the interviewee. Explain the purpose of your research, what it is, and what you will do with the information you gather. Address confidentiality issues, and get permission to record the interview. Explain what the interview process will be like and how long it will take.
- *Warm-up*. The aim is to help the respondent to relax. Emphasize that there are no wrong or right answers, and start with questions that are easy to answer and relatively impersonal.
- *Main body of the interview*. The main body of the interview consists of questions relating to the list of topics contained in the interview guide. These will consist of:
  - questions that introduce a topic (Would you please tell me about when your interest in the movement began? Have you ever . . . ?);

  - direct questions (Do you enjoy participating in these types of activities?); and
  - probing questions; these include *follow-up questions* (Could you say more about . . . ?), *specifying questions* (How did you react when you heard the news? What did you do after that?), and *interpreting questions* (Do you mean that your opinion of X changed?).
- *Cool-off*. When the main topics have been covered, the interviewer should begin to prepare for winding down the interview. In this 'cool-off' segment of the interview, the interviewer moves the discussion away from 'heavy' topics—those that have been most difficult to discuss or that have generated tension—and returns to more general questions. The aim is to end the interview on a lighter and brighter note.
- *Closure*. The interview will conclude with asking the respondent if they have anything they would like to add to what they have told you, and if there is anything they would like to ask you. You should not appear to be in a hurry to get away. Conclude the interview by offering the respondent sincere thanks for taking the time to talk with you.

Conducting a face-to-face interview

Interviewing skills

Interviewing requires skill. A good interviewer (1) puts the interviewee at ease without becoming overly familiar; (2) tries to minimize the social distance between themselves and the interviewee by dressing in a culturally acceptable and simple style; (3) is at ease and never in a hurry; (4) listens to answers, shows interest in what the interviewee says, and never shows any disapproval of the information received during the interview; (5) is sensitive to the respondent's mood, body language, time constraints, and different cultural norms; (6) probes and cross-checks in a thorough but sensitive manner; (7) makes sure the environment is free of noise and other people.

Recording the interview

If an interviewee hesitates or refuses to agree to audio recording, the only solution is to take notes. Usually, interviewees forget quickly that the recorder is on, but if they appear disturbed despite their consent that it be used, it should be stopped. However, even if you are using a tape recorder (or, more likely, a dictaphone), you should take notes of important non-verbal events or observations. You should take notes in a discreet way without interrupting the flow of conversation.

Keeping control over the interview

The interviewer has to ensure that the interview begins well, and introduce the research to interviewees in a clear and comprehensive way. To ensure that the information provided by different respondents will be comparable, the interviewer must make sure that all topics are adequately covered. Responses that go on too long or go off track should be politely stopped or steered back on track. In this case the interviewer might say something like: 'Thank you, this is interesting. Do you mind if I go back to the previous question? This (specify) is not yet fully clear to me.'

Probing

Open-ended questions tend to be very general (e.g. 'What do you think about . . . ? Why do you feel that way?'). Consequently, respondents may answer in too general a way or use adjectives that are too general to convey a clear response. In addition to being unclear or incomplete, responses might be irrelevant to the question, offer no answer at all, or appear untruthful. Whenever a response is inadequate or unsatisfactory, the interviewer can use probing. Probing consists of asking questions to encourage further conversation, without influencing the answer. For instance, when a response is *unclear* or *incomplete*, the interviewer might gain clarity or additional information by asking for a more specific response, or an explanation of a term. For instance, the interviewer might ask: 'Could you elaborate a bit?', 'Could you mention more possibilities?', 'Why do you think so?', 'Why do you feel that way?', or 'Would you say some more about that?' Or, if the respondent has started to give an example but did not finish, the interviewer might probe with questions such as: 'When was that?', 'Why did you do that?', 'How often does that happen?' If the response is *irrelevant* to the question, the interviewer might say something like: 'That is not exactly what I meant to ask about' and repeat the question, slightly elaborated.

An answer might be *non-responsive*, either because the respondent did not understand the question or because the question touched on sensitive information that the respondent hesitated to divulge. In case of non-response, the interviewer should repeat or rephrase the question without suggesting a response. If it appears that the question was not understood, the interviewer might say: 'Perhaps I wasn't really clear in the way I asked the question. What I meant to ask was …', followed by the question, phrased slightly more elaborately. If the respondent seems hesitant to answer, then the interviewer can stress again that the information will remain in confidence, and then repeat the question. If the respondent still objects to continuing, the interviewer can offer to skip the question. Sometimes an interviewer might suspect a response is *not truthful* because various parts of an answer contradict each other or because the respondent seems concerned to give the 'desirable' answer. The interviewer should communicate, not suspicion, but an interest in clarifying the situation, and say: 'I didn't ask the question clearly enough', or 'I didn't understand the answer'. They then might say: 'How should I interpret this?' Summarize the contradictory statements: 'You said X. You also said Y', and then wait for the respondent to continue.

Ending the interview

At the end, the interviewer should not only summarize the interview, but also respond to questions that came up during the interview, give advice (if necessary or asked for), and give an opportunity for further questions from the respondent. Such 'after-interview' discussions and questions should *always* be recorded, like all spontaneous information, because discussions can shed light on complicated, not yet fully clear issues from many preceding interviews.

Conducting online interviews

Online interviews can be either synchronous or asynchronous. *Synchronous online interviews* resemble a traditional research interview in that they *take place in 'real time'*. The interviewer and interviewee are online simultaneously, in a virtual environment that allows

real-time communication between two or more users. An example is an internet chat room, in which questions and answers are posted, read, and answered in 'real time'. *Asynchronous interviews take place in non-real time*, for example using email. Asynchronous online interviewing can also be conducted via bulletin boards, discussion groups, or web/internet forums.

Email interviews

An email interview can be an effective form of interviewing. Many people prefer email interviews because it is not necessary for a synchronous time to be scheduled. It might be the *only* way to conduct an interview (1) with a busy (or reclusive) public figure, with whom it is not possible to schedule a time to meet face-to-face or to speak on the phone; or (2) if the inconvenience and expense of travelling to an interview site would be prohibitive.

In email interviews, you should send an initial note to introduce yourself and to inform the individual of your affiliation, what information you are seeking, and how the information will be used. Indicate the length of time and the general time frame in which you foresee the interview process taking place. Ask the individual to specify a time for you to send questions that will be convenient and that will permit a timely response. Agree with the interviewee on a follow-up exchange for clarifications or follow-up questions if the responses you receive are not clear or if you do not understand something that the interviewee writes. The researcher then sends out an email which contains the interview questions, either in the body of the email or as a Word attachment to the email. The participant responds to the interview questions, again either in the body of the email or in a Word document, and returns the completed answers to the researcher. Often the interview will take place over a period of time and questions are sent in stages so that the interviewee is not overwhelmed with a long list of questions at the start of the process.

Email is the most familiar mode of online interaction, and is relatively simple in terms of technological requirements. Email interviews avoid all the practical difficulties of recording equipment and transcribing machines, and they eliminate travel and transcription costs. They present opportunities to interview individuals based anywhere in the world; and they tend to reduce interviewer bias. They also give interviewees time to think about questions and to give considered responses. The ability to redraft and edit questions and answers allows for increased flexibility and deeper reflection. It may be, too, that people tend to be more open with others in cyberspace than in real-world communication. The lack of visual clues can encourage candid interchanges and open up conversation in directions which otherwise might be avoided. Individuals who are shy and reticent about speaking in face-to-face group interactions may find that the virtual environment encourages them to open up and engage in more candid exchanges.

There are, however, also disadvantages in conducting email interviews. The medium requires fairly simple questions, as there is less opportunity for the interviewer to probe or clarify. Greater motivation and interest will be required on the part of interviewees: they have to be willing to take the time and effort to sit in front of a computer screen and read text, think, type, and maintain a logical thread of answering. Respondents will not be able to express themselves at anything near the rate they communicate in speech, so less information can be exchanged within a given time frame. Moreover, while respondents may

feel better able to discuss sensitive issues in an online environment than in a face-to-face interview, they might also be less open if they have concerns about privacy or sense a lack of control over the use of personal data. The lack of time restriction and the slowing down of the communication process may lead, not to greater openness, but to lack of engagement or to the production of 'socially desirable' answers, rather than the more spontaneous responses typical of synchronous interviews (both online and face-to-face). The delayed interaction and the inability to spontaneously direct the flow of conversation, or to prompt and probe participants, may lead to a paucity of data compared to synchronous interviews. Finally, though email and other types of online interviewing provide researchers with opportunities to interview individuals based anywhere in the world, not everyone has access to the internet. Young, wealthy, and educated people tend to use computers more than the elderly, poor, and uneducated.

Interviewing via video link and web camera

Asynchronous online interviewing can also be conducted via video link or web camera. Conferencing software is available to facilitate online synchronous interviews in a chatroom-type environment. The software packages can be downloaded by the participants. They provide a large 'chat' window in which the dialogue is displayed. Beneath this is a smaller window where users type their text, and press return; seconds later the contribution is displayed, prefixed with their name.

Using synchronous chat differs in a number of important ways from more traditional interviewing. Missing are the visual, non-verbal cues which the interviewer would normally rely on in order to build rapport and gain the trust of the interviewee. The interviewer must, therefore, give thought to ways of establishing rapport with interviewees. Virtual interviews also do not permit the natural flow of conversation in the way that telephone or face-to-face interviews do. Consequently, the number of questions should be limited. The number of open-ended questions should be limited, as well. More than a few questions in total will make the interview begin to feel laborious. If more information is needed, the interviewer can ask interviewees if they are willing to do a follow-up phone interview. If you will be asking confidential or personal questions, you must consider how to handle privacy issues.

Make sure you understand the equipment and process before the interview and, in particular, the type of microphone and what adjustments may need to be made in using it. Test your equipment and, if using a webcam, practise using it before the interview. Get a Frequently Asked Questions (FAQ) sheet on how to use the equipment and respond to possible glitches that might arise. Choose an area with a neutral background, and make sure the areas around you are clear and orderly. You will want to speak directly to the camera so that it appears you are speaking to the interviewee. Position your materials in close proximity to the camera so that when you glance at your notes, you will not appear to be looking away from the camera. Consider how to eliminate all potentially distracting noises, as the microphone on a webcam can magnify the slightest sound. If using Skype, you may experience some time lag when you're talking. Make sure you talk clearly so your voice will be easily picked up by the microphone.

Focus groups, which we discuss later on in this chapter, can also be conducted online. Online focus groups can be conducted using **listserv**, which is one of the key

software applications for managing mailing lists. One advantage of this is that it eliminates the need to set up mutually convenient chat times. However, the interviewer does not communicate synchronously with participants. Respondents can post their replies at any time. This reduces the level of group interaction and the sense of immediacy. And the fact that the facilitator cannot always play an active role in moderating the discussion means that there is a greater likelihood of exchanges being unfocused or off-topic.

Elite interviews

Much political research is concerned with understanding political institutions and decision-making processes. Interviewing political elites is a key means of obtaining information about many aspects of these phenomena. Political elites are people 'who exercise disproportionately high influence on the outcome of events or policies in your research area' (Pierce 2008: 119). We can identify three broad purposes for which interviewing elites can be useful.

The first purpose is to obtain new information. By virtue of the positions they hold, elites may have access to information that might not otherwise be available to a researcher. However, a second purpose of elite interviews is to 'confirm the accuracy of information that has previously been collected from other sources'. As Oisin Tansey explains:

> When documents, memoirs, and secondary sources provide an initial overview of the events or issues under examination, interviews with key players can be used to corroborate the early findings. In this way, interviews contribute toward the research goal of triangulation, where collected data are cross-checked through multiple sources to increase the findings' robustness.
>
> (Tansey 2007: 766)

Finally, elite interviews can enable a researcher to make inferences about the beliefs or actions of a wider population of political elites. When analysts randomly select a sample of elites to interview from among a broader population, they can generalize from the findings of that sample to the wider group (see e.g. Aberbach and Rockman 2002).

One of the biggest challenges of conducting elite interviews is getting the individuals you select to agree to an interview. What you will need to do is to write to prospective interview subjects. This should be a brief letter 'on the most prestigious, non-inflammatory letterhead you have access to' (Aberbach and Rockman 2002: 674). The letter should state your purpose, but not in great detail. It should suggest, in a subtle way, why the interview would in some way benefit the prospective interviewee. Roger Pierce (2008: 120) lists a few reasons why elites might want to agree to an interview. These might include:

- 'you have persuaded them of the importance of your research and their potential contribution';
- they feel that their position obligates them to agree to interviews and that requests for interviews underline their status;
- they want an opportunity either to 'set the record straight' or to reflect on the policies, politics, or institutions with which they are or have been associated.

Your initial letter should also inform the recipient of your intention to follow up with a phone call to see if a time might be arranged to meet. You must then follow up the letter with a phone call. Prepare for an exchange with the appointments secretary, in which you will deliver a charming and persuasive account of your seriousness of purpose. Be persistent and 'insist firmly, but politely (and with a convincing explanation)' that no one but the targeted individual 'will do for the interview' (Aberbach and Rockman 2002: 674).

We haven't said anything about how to choose interview subjects. Roger Pierce suggests that you 'aim higher rather than low by approaching the "A-List elites" rather than lesser elites whom you may consider to be more approachable' (Pierce 2008: 121). But, as George and Bennett point out, often lower-level officials may be better interview sources, given their day-to-day involvement with political processes (George and Bennett 2005: 103). As with all research decisions, you must let your research question and hypothesis determine your choice of interview subjects. We would like to point out, however, that you should not hesitate to pursue 'A-listers' that you feel would contribute to your research. A-listers are always happy to receive requests for their time and attention; they may, in fact, feel uneasy whenever such requests begin to thin out.

Let's move on. Once you have an interview scheduled, you must decide on a format. Elite interviews tend to be semi-structured and use open-ended, rather than closed questions. But there are a number of factors you should consider in deciding what type of question to use. One consideration is the degree of prior research on the subject of concern. The more that is known, the easier it is to use closed questions. Conversely, the more uncharted the terrain, the more likely you will want to use exploratory, open-ended questions. A second consideration is 'response validity'. Open-ended questions increase the validity of the responses because they provide a greater opportunity for respondents to organize their answers within their own frameworks. The third major consideration is the receptivity of respondents. 'Elites especially—but other highly educated people as well—do not like being put in the strait-jacket of closed-ended questions. They prefer to articulate their views, explaining why they think what they think.' Reflecting on their experience in interviewing elites, Joel Aberbach and Bert Rockman noted that closed questions 'often elicited questions in return about why we used the response categories we used or why we framed the questions the way we did' (Aberbach and Rockman 2002: 674).

Though elite interviews often rely on open-ended questions, a combination of open-ended and closed questions might provide researchers with better opportunities to obtain the information they are seeking. For example, Sharon Rivera and her colleagues used a combination of open-ended and closed questions to interview Russian elites. They started out with five open-ended questions, then followed with a couple of closed questions, and then alternated between open-ended and closed questions for the remainder of the interview. With this sequencing, they were able to realize a number of advantages. By starting with open-ended questions, they found it easier to elicit answers to the closed questions. As they describe it:

> We had demonstrated respect for the complexity of their views through the open-ended questions and thus had 'earned' the right to ask questions posed exclusively from our frame of reference. Also, the closed-ended questions probably allowed respondents to recover a bit from the more demanding open-ended question format.
>
> (Rivera, Kozyreva, and Sarovskii 2002: 686)

In addition to, or as part of the process of, preparing questions for the interview, you must thoroughly research the subject. Make sure you know as much about them as possible. You should also check online for images so that you will be able to recognize your interviewee by sight. Arrive at the interview with your questionnaire, and/or **interview schedule** or guide, and an audio recorder. Whether or not you record the interview, you will want to take notes (to record non-verbal cues, as well as your own reflections); but you should not make the note-taking obvious, or let it deter you from maintaining eye contact. In conducting the interview, you will need to give space for the respondent to open up new areas, while making sure to cover the topics you feel you need to address. End the interview ten minutes early. The most revealing information is expressed by interviewees when the interview appears to have been formally ended. After the interview, type up your notes and transcribe your recording as soon as possible.

Expert interviews and surveys

Expert interviews and surveys might be thought of as a type of 'elite interview'. But the type of data researchers seek to collect through expert interviews and surveys will usually differ quite substantially from the data they might obtain through the use of elite interviews. Expert surveys are widely used in comparative research to measure a wide variety of different political phenomena, and can be analysed both qualitatively and quantitatively.

In an expert interview or survey, the researcher identifies individuals with specialized knowledge or expertise relating to a particular issue and asks them a common set of questions, either directly in face-to-face interviews, or through the administration of a postal survey. Rather than probing to discover what selected individuals think, and why they think what they think, the aim of expert interviews is to elicit specific information from individuals with specialized knowledge or expertise on a particular issue. Expert surveys perform the same function. Expert surveys enable researchers to summarize the consensus judgement of experts on a given issue and do so in a systematic way (Benoit and Laver 2006: 9).

Of course, the validity of expert-opinion data will depend on the quality of the experts. Expert judgements are valid and reliable in so far as the experts are well-informed about the subject in question. Some studies use a large pool of experts; others rely on a selected few. The use of multiple experts increases the validity of the data (Dorussen, Lenz, and Blavoukos 2005: 317). But in some instances knowledge may be limited to a handful of observers or participants, 'either because only a few persons are privy to the relevant information (e.g. government positions in behind-closed-doors negotiations) or because only a few persons have the relevant expertise' (Hooghe et al. 2009: 4).

Expert-opinion data has been used widely in studies of the European Union. An example is the Domestic Structures and European Integration (DOSEI) project. The purpose of the project was to determine the position of the EU members on the draft Constitution for the European Union. As part of the DOSEI project, 'country experts were asked to assess the national position on the draft Constitution. A total of seventy-seven experts were interviewed, varying from one to six experts for any particular political actor' (Dorussen, Lenz, and Blavoukos 2005: 316). The project collected a wealth of information on all the relevant policy positions of all the main actors with respect to the draft European Constitution.

Expert surveys have also been used to identify the policy preferences of political parties. What these surveys seek to do is to obtain a consensus of specialists about particular

policy positions in order to identify where parties stand on national issues. In a study that revealed the potential gains that could be produced from using expert surveys for this purpose, Francis Castles and Peter Mair (1984) used a postal survey to ask experts in sixteen countries to locate parties in their 'own' countries on a ten-point scale between Left and Right. Kenneth Benoit and Michael Laver (2006) conducted a similar survey covering forty-seven countries. The 'experts' they surveyed were typically academic specialists in the political parties and electoral politics of their 'own' countries. Benoit and Laver defined four substantive policy dimensions as a basis for comparing policy scales among surveyed countries. For each policy dimension (economic policy, social policy, the decentralization of decision-making, and environmental policy), they used a scale running from 1 to 20, with the lower position indicating the typically 'left' position and the higher value the traditionally 'right' position.

Expert surveys are also used to collect a wide variety of data in comparative research on topics such as democracy, corruption, human rights, and political repression, to name but a few (see Chapter 9). Often hard data on these issues is either unreliable or unavailable. For example, very few governments keep up-to-date records on how many political prisoners they torture. Nor is there much in the way of official data on how corrupt business practices are. To get round this problem researchers often rely on evaluations of different political phenomena by experts in the relevant field. So, who better to assess how corrupt business practices are in different countries than business people who frequently travel around the world? Transparency International has now been carrying out expert surveys on this issue for a number of years. It asks business people a series of questions about business practices and corruption in different countries. The data do not therefore strictly measure how corrupt a country is. Rather they measure perceptions of corruption. But the more business people agree on whether or not a country is corrupt, the higher we can regard the validity of the measure. We discuss this issue of intercoder reliability in more detail in Chapter 14, but the principles are much the same in expert interviews.

An example helps to illustrate this point. Expert surveys are frequently used to measure the level of democracy in a country. Two of the most widely used sources of data on this are Freedom House and Polity IV. Both draw on expert surveys (to differing degrees). Experts (defined as country specialists) are asked to evaluate the extent to which specific countries meet different criteria. For example, one question on the political environment asks, 'To what extent are media outlets' news and information content determined by the government or a particular partisan interest? (0–10 points).' Although there is a certain amount of subjectivity inherent in this approach, ideally, if all the experts are unbiased and well-qualified to answer this question, then they should all agree and give the same rating. Under these circumstances intercoder reliability is high, and we can think that perceptions are a valid measure of actual experience.

Focus groups

The focus group can be considered a form of group interview. However, it differs from a standard group interview in two ways. First, it involves a group of interviewees who are selected on the basis of their having had, or not had, some particular experience, or having a particular attribute. Second, rather than asking subjects to respond to the same question, as in a group

interview, the focus group session is organized to address a theme or topic. Alan Bryman characterizes a focus group as combining a group interview with a focused interview (2015: 502).

Focus groups are a form of interviewing involving a group of six to ten people who meet in a conference-room-like setting, and are interviewed together in a flexible and exploratory group discussion format. Members of the focus group are selected because they share some attribute or are related to some phenomenon of interest. They might be chosen, for instance, because they are all between the ages of 35 and 50, or all have a child currently serving in the army and based overseas, or are all female police officers. The people selected to form a group can be unknown to each other or already form some sort of pre-existing or natural group—people who know each other through clubs or social centres, who work together, or are classmates.

Rather than acting as an interviewer, the researcher performs the role of moderator or facilitator. The moderator initiates the session and moderates the ensuing discussion. The facilitator generally comes equipped with a relatively small number of questions and a willingness to relinquish a larger degree of control over the proceedings than would normally be the case in an individual interview, and to allow the discussion to develop and move in different directions with a certain degree of spontaneity. The emphasis is on interactions between participants rather than between the researcher and participants. The purpose is to explore people's ideas in a public setting so that the interviewer can observe how they react to each other's ideas and how their opinions are formed.

Why do focus group interviews rather than individual interviews? Interviewing individuals, face-to-face, or over the telephone, or through mailed questionnaires helps researchers learn about how people feel. The focus group is a good technique for exploring social interactions, and how these shape attitudes and opinions. Exchanges among participants can lead to far more probing and reflection than is possible in individual interviews or questionnaires, and may provide more robust and revealing responses to the issues which are the subject of the focus group. The discussion can also bring into focus dimensions of an issue not previously considered by the interviewer. The unique contribution of a focus group among the array of interview methods is that it enables a researcher to study 'the ways in which individuals collectively make sense of a phenomenon and construct meanings around it' (Bryman 2015: 348).

In Chapter 2 we considered how understandings emerge as a result of interaction and discussion as individual participants contrast and fit their individual views with those of others. The understandings that emerge from interaction among individuals represent, not just the sum of their separate views, but something 'novel, not possessed by the individuals taken in isolation and that would not have emerged in the absence of their interaction'. Interactive relations between individuals can produce properties that are separate and distinguishable from the individuals themselves.

When a group of people with similar interests discuss an issue together, they are likely to produce richer insights and a wider range of information than individual responses obtained privately. One person's comment is likely to provoke a reaction from the other participants and generate more views. Since members of focus groups are selected for some shared attribute or experience, they are likely to feel more comfortable expressing their ideas and feelings; and, because they are not required to answer specific questions, respondents may be more spontaneous, and accurate, when expressing their views. However, a major determinant of focus group success and of the quality of their results is the skill of the moderator. Also, focus groups are not representative of the general population; and the

unstructured nature of the responses in focus group discussions makes coding, analysis, and interpretation difficult. Hence, the results of focus group discussions should be treated as exploratory, rather than conclusive.

Organizing and conducting a focus group session

Organizing and conducting a focus group session involves the following steps:

1. *Identify the major objective of the focus group session*. Just as is the case with questionnaires or interview guides, focus groups must be designed to produce data relevant to a hypothesis.
2. *Develop questions*. A session should last between sixty and ninety minutes. This is time enough to allow the moderator to ask at most five or six questions. Make sure these are formulated so as to elicit the information you want to be able to gather during the session.
3. *Select the membership.* Focus groups are usually conducted with six to ten members who have some similar attribute or experience. Select members who are likely to be participative and reflective, and who don't know each other.
4. *Plan the session*. Hold the session in a conference room, with participants sitting in a circle or around a seminar table so that everyone can see everyone else. Provide name tags for members, and refreshments. The *agenda* would normally look something like this: (1) welcome the participants; (2) review the agenda; (3) explain the goal of the meeting, and discuss the ground rules; (4) have group members introduce themselves; (5) questions and answers; (6) wrap up.
5. *Running the session*. The moderator will normally begin a session by introducing themselves, the purpose of the research, the format of the session, and some ground rules, such as allowing everyone to speak and only one person to speak at a time. The moderator will also want to ask participants to fill out a form with some basic information about themselves: age, occupation, etc. In addition to an information sheet, participants should also be a given a consent form to sign before beginning the focus group.

All members should participate as much as possible. If one or two people are dominating the meeting, then call on others. Tell those individuals that you want to hear from the others. Don't look at the individuals when you ask a question. Raise your hand as if to say 'stop' when the individual tries to talk, and look at someone else. You might say something like, 'Let's have someone else go first'. If the participant continues to dominate the session, you might suggest using a round-table approach, including going in one direction around the table, giving each person a minute to answer the question. If the domination persists, note it to the group and ask for ideas about how the participation can be increased.

Don't be abrupt with a dominating participant, as it may discourage participation from others. And if you feel your efforts to broaden the discussion have led to a withdrawal of participation by the dominating individual, try to re-establish rapport. If nothing else works, take a break and privately either discuss your concerns with the individual or ask the individual to leave. The discussion should remain focused, maintain momentum, generate useful information, and achieve closure on questions. After each question is answered, the moderator should reflect back a faithful summary of what was said.

6. *Record the session with either an audio or audio-video recorder*. The focus group session should be recorded, using microphones strong enough to pick up voices around the table or circle. It will be important to have a record, not just of what was said, but of how it was said and who said it. So there should be someone taking notes as well.

The Research in Action Box 12.1 provides an example of research using focus groups data which makes an important contribution to current debates on the impact of anti-terrorism legislation and policy measures on citizenship.

Research in Action 12.1 The Impact of Anti-Terrorism Policy on Citizenship

Between 9/11 and 2012, there was 'a dramatic expansion of the UK's domestic anti-terrorism architecture' (Jarvis and Lister 2013: 657). Major legislative Acts expanded the government's anti-terrorism powers by bringing into being the power of detention without charge for foreign nationals (this has now been repealed), increased pre-charge detention periods, a control orders regime, and a host of new 'community resilience measures'. What have been the consequences of these policies for public life?

As Lee Jarvis and Michael Lister point out, 'much of the academic literature exploring anti-terrorism and citizenship argues that legislation and policy measures introduced under the "war on terror" have brought about a diminishment of citizenship' (2013: 658). Recent research has highlighted the '*particular* difficulties experienced' by Britain's Muslim communities as, for instance, 'the disproportionate increase in uses of stop and search powers upon Asian individuals'; but it also notes that the effects of these policies (intended and otherwise) ripple more widely' (2013: 660) as, for instance, the 'civic obligation to contribute to, and participate in, the monitoring of others' (2013: 661).

Missing from the literature, however, is 'sustained engagement with these dynamics as experienced and/or articulated by' citizens themselves (2013: 661). Citizenship, the authors argue, is as much a practice. 'a lived, enacted experience', as a formal legal status. Consequently, to understand the impact of these policies requires a richer and more detailed understanding of 'how citizens themselves perceive and understand the effects of anti-terrorism architectures on their lives – as citizens' (2013: 661). 'Do people within the UK feel that their citizenship has been compromised by anti-terrorism legislation? Are there different responses to this question based on ethnic or other forms of identification?' (2013: 662).

To answer these questions the authors organized a set of focus groups in which they sought insight into the perceptions of different ethnic groups (black, white, Asian) and within these groups, individuals living in metropolitan and non-metropolitan areas. Participants were asked a series of open-ended questions, and transcripts of their discussions were later analysed. Four themes emerged from this analysis: (1) freedoms and rights, (2) the ability and desire of individuals to participate in the public sphere, (3) obligations and duties, (4) the relation of individuals to British identity. Further analysis was undertaken to determine how these multiple dimensions of citizenship interacted in the lives of citizens from different population groups.

The authors found that white participants, though concerned about these measures, viewed them 'primarily in the abstract' (2013: 667). White and ethnic minority participants had strikingly different views in relation to their sense of belonging, perception of duties and rights, and political engagement. While anti-terrorism measures, it appeared, had had only a limited impact on the former group, the latter group expressed a declining sense of belonging and dampened political engagement; and this, and its impact on their perceptions of duties and rights, produced 'a vicious cycle of declining citizenship' (2013: 672). Based on these findings, the authors concluded that anti-terrorism measures may be contributing to an experience of differential or 'disconnected' citizenship in the UK, which, they suggest, may work to weaken British citizenship overall.

Analysing interview data

Analysis of interview data entails three main steps: **data reduction**, coding, and drawing conclusions. Indeed these steps of analysis are common with all types of data, though how we go about doing it in practice varies somewhat according to the type of data we have collected (e.g. see the discussion of textual analysis in Chapter 14).

Data reduction

'Data reduction refers to the process of selecting, focusing, simplifying, abstracting, and transforming the data that appear in written up field notes or transcriptions' (Miles and Huberman 1994: 10). Data reduction can be achieved by noting redundancies in the data and discarding all but the most interesting and compelling statements concerning a particular issue or theme. Not only do the data need to be condensed for the sake of manageability, they also have to be transformed so they can be made intelligible in terms of the issues being addressed.

The data from unstructured interviews is not as easy to code or quantify as data obtained from surveys. Great care needs to be taken in developing code frames to record the different types of answers, and attention also needs to be paid to how interviewers' questions or probes vary across different respondents. Since there is often substantial variation in how questions are asked, the data cannot be tracked as easily over time, and the results of data analysis are not as easily compared, since in effect what is often being tracked or compared is answers to different questions. The more structured a question, then the easier it will be to analyse. Data for focus groups are difficult to analyse—particularly if the analysis is concerned with both the content of what was said and the patterns of interaction. Open questions permit free responses, which are recorded or written down as far as possible in the respondents' own words. This mass of data has to be organized and somehow meaningfully reduced or reconfigured. Box 12.3 gives a snapshot of how the analysis and results obtained from semi-structured and unstructured, and structured interview data differ.

The first step in data reduction is to transcribe the tapes or notes of the interviews. As part of this process, you should develop a 'profile' of each interviewee: a summary of the interviewee's background, experience, and opinions. Whenever possible, interviewees' opinions should appear in these profiles in their own words, to avoid the possibility that rewriting them may change their substance or emphasis.

BOX 12.3 Different Processes of Data Analysis

Structured interviews	Unstructured interviews
Analysis is done using standardized statistics and procedures	Data is analysed by systematically organizing and interpreting information using categories, themes, and motifs that identify patterns and relationships
Results tend to summarize patterns of similarities, variability, size, direction, and/or significance of any differences between specific groups	Results are in-depth explanations for patterns of behaviour

Recording and transcribing interviews results in an unwieldy amount of verbal data. One of the central tasks of semi-structured and unstructured interview data analysis is to reduce the amount of data to a more manageable level by identifying and extracting the most important, meaningful, and interesting parts of the interview text. This should begin as a process of discovering what's in the material, rather than starting out with definite hypotheses in mind. When the transcripts have been completed, read them first as a whole to note your general impressions. Read carefully through each interview without taking notes. When you finish reading you might wish to write down a few general notes or impressions.

Next, review the interview transcripts again, this time looking for very specific things—for similarities and differences, and patterns and thematic connections in the data. Make marginal notes, highlighting key words and themes. For focus groups, make a note of major opinions and attitudes that are expressed by the groups. Underline or circle those items that you think will prove to be the most important or meaningful. This is the start of producing a list of terms that will subsequently be used for coding. Codes emerge from a process of refining the marginal notes you make on transcripts.

Coding

Coding means breaking down interview material and assigning data to different categories according to the variable to which they relate. In coding interview data you name and categorize phenomena through close examination of data. Closed questions are pre-coded. Open questions will require a coding frame—a list of categories and codes for each question used to analyse responses. Eventually, you will perhaps want to reduce the number of code words and categories of information; but as you begin this process you should not worry about placing too many initial restrictions on yourself.

Consider whether you are using two or more words or phrases to describe the same thing; or whether an item might be coded in more than one way. Some responses may refer to more than one issue or idea. These should be given more than one code word so that the response can be included in each relevant area. Consider, too, whether your codes might be replaced with concepts and categories in the literature relating to your research question and hypothesis. Look to see if certain codes tend to be associated with other codes, and whether you can perhaps code for these connections.

When you code the transcripts, you mark sections of the transcript in a way that indicates what the interviewee or focus group participant is talking about. For instance, every time a participant mentions a 'news report', the researcher marks the section to indicate this. Using code words will make this faster, e.g. NEWSREP (i.e. NEWS REPORT). So, in the end a transcript will have a list of code words running down the side of the page. This makes it easier to identify sections of interest later on, as all that will be needed when looking at the issue of XYZ is to look down the transcripts and take all the responses marked XYZ.

The information provided by open questions needs to be coded in order to make sense of the answers. This can be very time-consuming. Many responses may be uncodable; answers may be irrelevant or off-topic, and some groups of people may be more likely to respond than others, particularly those who are more educated or more interested in the topic. Each coherent interview segment should be given a code number to indicate the concept, category, theme, or argument it relates to. These segments can then be copied and separated

from the transcripts (with either real or electronic scissors) and put back together thematically to serve as the raw material for an analysis of the general findings of the study and their implications for policy- or decision-making.

Throughout this process it is important to reflect on what significance these data have for your research question and hypothesis. The researcher must make choices about which aspects of the data should be emphasized, minimized, or set aside completely. At all stages of the research process, the researcher will be concerned to probe the generalizability, reliability (the consistency of findings/results), and validity (if the study in fact investigates what was intended).

Traditionally, coding involved gathering together those portions of a text belonging to a given label or name. Researchers used different coloured highlighters for each code and then cut and pasted together similarly coloured fragments using scissors and paste. Today, there is computer software that performs this task. Coding of data can be done using one of the computer-based analysis program packages (e.g. NUD*IST, NVivo, or Atlas.ti). However, whether computer software or highlighters are used, the process remains the same. It is the researcher who defines and names the categories of data. The result of both coding and analysis depends exclusively upon the researcher's interpretation of meanings hidden in the data.

Using the research question and hypothesis as a guide, every line, paragraph, or other section of text is coded for relevant themes. As themes are developed, the researcher assigns a working definition to each code. That way, in going through the transcripts, the definition is continually being challenged, and sometimes new codes must be developed because the properties do not fit the text. Also, codes that are rarely used are dismissed and some categories are broadened to accommodate the lost code. Important to note is that this type of analysis is not linear, but circular. Constant comparison (see Glaser and Strauss 1967) means that the researcher must continually compare the categories and codes of new transcripts with the existing categories and codes in order to more fully develop the properties of the overarching categories for the individual codes. This process is ongoing until saturation is reached: that is, no new codes or categories emerge and coding more transcripts would only produce a repetition of themes.

Be aware that extracting and coding fragments will remove them from the context within which they appeared and disrupt or lose the narrative flow of what was said (Bryman 2015: 411).

Analysis

Once sequences of text are marked with codes (coding), all the sequences of text marked with each code are collected together (retrieving). The analysis should then connect up the codes to each other in order to bring into focus a web of meanings.

Drawing conclusions involves stepping back to consider what the analysed data mean and to assess their implications for the questions at hand. Verification, integrally linked to conclusion drawing, entails revisiting the data as many times as necessary to cross-check or verify these emergent conclusions, to validate findings by cross-checking with other questions and information from other respondents. 'The meanings emerging from the data have to be tested for their plausibility, their sturdiness, their "confirmability"—that is, their validity' (Miles and Huberman 1994: 11). Validity means something different in this context than it does in quantitative evaluation, where it is a technical term that refers quite specifically to whether a given construct measures what it purports to measure. Here validity

encompasses a much broader concern for whether the conclusions being drawn from the data are credible, defensible, warranted, and able to withstand alternative explanations.

Folkestad (2008) is interested in exploring Euroscepticism among various political parties in Nordic countries and Central Eastern Europe. He decided that conducting qualitative interviews with elites in political parties might provide him with insight into this phenomenon. To explore the content of Euroscepticism among the respondents that he interviewed, he formulated conceptual categories concerning the trajectory of the EU, drawn from the categories that Kopecky and Mudde (2002) used to conceptualize support for and opposition to European integration.

Folkestad reduced the material from the interview into three categories:

> right and wrong trajectory of the EU, and a more ontological category, 'Nation-state'. Whereas the two first categories show the answer the respondent is giving to the question (right and wrong directions of the EU), the last category tells us something about what kind of political organization/polity they view as the most 'proper'.
>
> (Folkestad 2008: 7)

The data reduction process is shown in Figure 12.1.

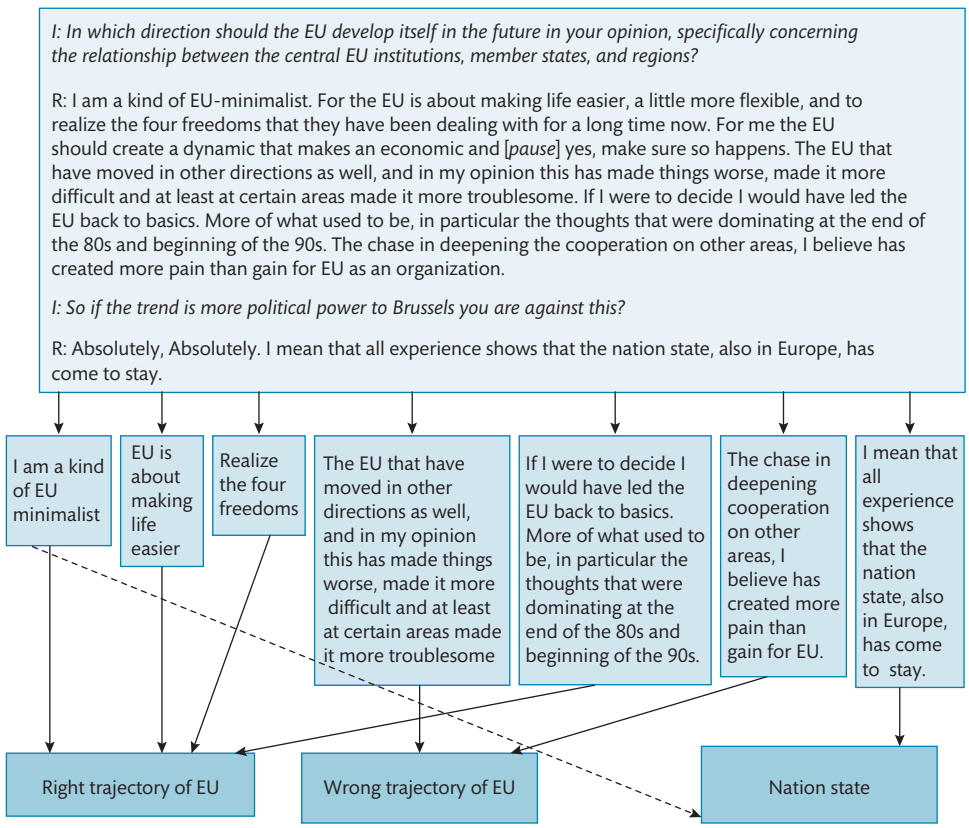


Figure 12.1 An example of data reduction

Source: Adapted with permission from B. Folkestad (2008). Analysing Interview Data: Possibilities and Challenges. *Eurosphere Working Paper No.13*. Bergen: Eurosphere Consortium. eurospheres.org.

Box 12.4 reviews some key terms introduced in this chapter.

BOX 12.4 Key Terms

Structured interviews	A form of interview that consists of a standardized set of closed and shorter or simpler questions that are asked in a standardized manner and sequence and, thus, are better for making comparisons.
Unstructured interviews	A type of interview that uses open, and perhaps lengthier and more complex questions, which might vary in the way and order in which they are asked, and which allow for more in-depth probing and flexibility, than structured interviews.
Semi-structured interviews	A form of interview in which the interviewer uses a combination of structured questions (to obtain factual information) and unstructured questions (to probe deeper into people's experiences).
Interview guide	In conducting unstructured or semi-structured interviews, the interviewer will generally prepare an interview guide: a set of questions that provide a list of talking points or areas to be covered.
Elite interviews	A method of obtaining, from political elites, information that might not otherwise be available to a researcher, or that can confirm the accuracy of information that has previously been collected from other sources.
Expert interviews	A method of eliciting specific information from individuals with specialized knowledge or expertise on a particular issue.
Focus groups	A type of interview in which the researcher meets with a group of people, selected because they are related to some phenomenon of interest, and facilitates an organized discussion related to something the participants have experience of or beliefs about.
Data reduction	The process of condensing (selecting, focusing, simplifying) the data that appear in written up field notes or transcriptions for the sake of manageability and so they can be made intelligible in terms of the issues being addressed.
Coding	The process of breaking down interview material and assigning data to different categories according to the variable to which they relate.
Validity	In the analysis of interview data, 'validity' refers to whether the conclusions being drawn from the data are credible, defensible, warranted, and able to withstand alternative explanations.

Conclusions

As with any method of data collection, the use of interviews or focus groups should be appropriate both to the research question and to the hypothesis you are investigating. You choose interviewing as a means of data collection because you are confident that the data that you collect through this means will provide a convincing test of your

hypothesis, and will enable you to draw logical conclusions about your hypothesis. Questions should be designed so that the responses actually measure what they claim to measure.

As we discussed in Chapter 6, it is a good idea to use multiple sources of data and methods of data collection whenever possible. Doing this enables you to approach a research problem from different angles. This is called 'triangulation'. Triangulation of evidence increases the reliability of the data and the process of gathering it. In the context of data collection, triangulation serves to corroborate the data gathered from other sources: the use of different data sources can enable researchers to cross-check findings. Triangulation yields more complete data and results in more credible findings; and it also enables researchers to find agreement between different perspectives. It might involve the use of different research methods to study a single research problem—a comparison of the results of these different methods can enable a researcher to identify whether differences are due to biases in one or another of these methods. It might also involve the use of different researchers to study the same research problem with the same people and with the same methods: if the same results are discovered, the findings and interpretations have stronger validity because of this corroboration.

In Chapter 6 we discussed, in general terms, the ethical principles to which all political research should conform. These principles have particular relevance to research that involves conducting interviews and organizing and leading focus groups. Thus, to conclude this review of interviews and focus groups, we want to draw your attention to the particular ethical issues that use of these research methods raise (see Box 12.5). We also provide a template for (1) informing your interview subjects and focus group participants about the procedures that you will be following, and (2) securing their consent to participate (see Box 12.6).

BOX 12.5 Ethical Issues in Interviewing and Focus Groups

Always be honest about who you are, the purpose of your research, why you want to talk to your interviewee or focus group participants, what you will do with the information, and the level of confidentiality you can promise. Explain what the interview or focus group process will be like and how long it will take. Supply all participants with a written version of the procedures that will be followed. In addition, you should:

- tell participants that their participation is voluntary;
- obtain written consent for participation;
- obtain permission to record the interview or focus group;
- tell participants that they may withdraw from the research at any time and for any reason;
- tell participants that their data will be treated with full confidentiality and that, if published, the date will not be identifiable as theirs.

BOX 12.6 Checklist for Information Sheet and Consent Form

1. In the information sheet

[Department], [University] (understandable to lay person)
Include your own name (with contact details) and supervisor's name (if applicable)
Use lay terminology to explain study—aims (why doing it) and what will happen if decide to take part
Where study is taking place
Participation is entirely voluntary
Participation is anonymous and confidential (only seen by myself and supervisor (for student researchers only)) (NB only include that it will not be shown to teachers and carers where applicable—often not relevant)
Can decide not to answer any question if you prefer not to
Can withdraw at any time without giving a reason (and without affecting education or care if applicable)
Your signed consent form will be stored separately from the responses you provide
If you decide not to participate, it will not affect your education or care (only mention this if applicable)
NB: You may retain this information sheet for reference and contact us with any queries

2. Consent form

(Please note that researcher can use tick box, initial box, or delete YES/NO format, but whichever you use, there must be a mark in response to each of the 'please indicate' items):
Name of study and researcher

	Please indicate
I have read the information sheet about this study	(YES/NO)
I have had the opportunity to ask questions	(YES/NO)
I have received satisfactory answers to any questions	(YES/NO)
I understand that I am free to withdraw from the study at any time, without giving a reason	(YES/NO)
I agree to participate in this study	(YES/NO)

Signed … … … … … … … … … ….

Name … … … … … … … … … … Date … … … … … … … … … … …

NB: This consent form will be stored separately from the responses you provide.
Please note: There should be *no data collected on the consent form* as this will be stored separately from data.

Source: Royal Holloway, University of London Ethics Committee, 25 March 2014

Questions

1. What are the advantages and disadvantages of semi-structured interviews compared to structured interviews and surveys?
2. How can interview data be used to confirm or disconfirm a hypothesis or argument?
3. What considerations would guide your choice of what type of interview to conduct?

4. What considerations need to be taken into account in the formulation of interview questions?
5. What behavioural observation and interpretation skills are needed to assist the interview process?
6. What are the benefits and drawbacks of using focus groups to generate data?

Guide to Further Reading

Aberbach, J., J. Chesney, and B. Rockman (1975), 'Exploring Elite Political Attitudes: Some Methodological Lessons', *Political Methodology* 2(1975): 1–27.
Uses data from interviews with American political elites to illustrate open-ended interviewing. Discussion on coding techniques for interview data analysis.

Barbour, R. (2018), *Doing Focus Groups*, 2nd edition (London: Sage).
A fine introduction to the different purposes for which focus groups are used in research, the different forms they can take, research design issues, the planning and running of focus groups, and the nature and status of the date that are generated.

Denzin, N. K. and Y. S. Lincoln (eds) (1994), *Handbook of Qualitative Research* (Thousand Oaks, CA: Sage Publications).
This volume consists of thirty-six chapters on qualitative methods in social research. It covers historical and philosophical perspectives, as well as detailed research methods. Extensive coverage is given to data collection and data analysis, and to the 'art of interpretation' of findings obtained through qualitative research.

Gubrium, J. F. and J. A. Holstein (eds) (2002), *Handbook of Interview Research: Context and Method* (Thousand Oaks, CA: Sage Publications).
A comprehensive handbook covering virtually all forms of interviewing.

Hammer, D. and A. Wildavsky (1993), 'The Open-Ended, Semistructured Interview: An (Almost) Operational Guide', in A. Wildavsky (ed.), *Craftways: On the Organization of Scholarly Work* (New Brunswick, NJ: Transaction Publishers), 57–101.
Examines the procedures associated with semi-structured interviewing and provides useful discussion about the relationship between the researcher and the respondent.

Krueger, R. A. (1998), *Moderating Focus Groups* (Thousand Oaks, CA: Sage Publications).
A detailed guide to leading focus groups, with much useful advice and helpful strategies for handling a range of issues.

Kvale, S. (1996), *InterViews: An Introduction to Qualitative Research Interviewing* (Thousand Oaks, CA: Sage Publications).
Covers all aspects of the interview process, including the theoretical and philosophical foundations of interviewing, and the influence those foundations have on the process, the content, and the analysis of qualitative interviews.

MacNaughton, P. and M. Jacobs (1997), 'Public Identification with Sustainable Development: Investigating Cultural Barriers to Participation', *Global Environmental Change* 7: 5–26.
An instructive example of the use of focus groups to address a research question, in this case to investigate the role of 'culture' in public attitudes about sustainable development initiatives.

McLellan, E., K. M. MacQueen, and J. L. Neidig (2003), 'Beyond the Qualitative Interview: Data Preparation and Transcription', *Field Methods* 15 (February): 63–84.
Outlines the consequences of inappropriate or inadequate preparation of transcripts from recordings and offers practical considerations that can help researchers systematically organize and analyse textual data.

Morgan, D. (1997), *Focus Groups as Qualitative Research*, 2nd edition (Newbury Park, CA: Sage Publications).
Compares participant observation and individual interviews; strengths and weaknesses; uses of focus groups.

O'Connor, H., C. Madge, R. Shaw, and J. Wellens (2008), 'Internet-based Interviewing', in N. Fielding, R. M. Lee, and G. Blank (eds), *The SAGE Handbook of Online Research Methods* (London: Routledge), 271–89.
This introduces the potential and limitations of internet-based interviewing, different types of online interview and examples of each, interview design, ethical issues, and suggestions about different types of software used for online interviews.

Sudman, S. and N. Bradburn (2004), *Asking Questions: A Practical Guide to Questionnaires*, revised edition (San Francisco, CA: Jossey-Bass).
A comprehensive guide to how to ask all kinds of questions, and how to draft and craft a questionnaire.

'Symposium: Interview Methods in Political Science' (2002), *PS: Political Science and Politics* 35(4) (December): 663–88. B. L. Leech, 'Asking Questions: Techniques for Semistructured Interviews' (pp. 665–8); K. Goldstein, 'Getting in the Door: Sampling and Completing Elite Interviews' (pp. 669–72); J. Aberbach and B. Rockman, 'Conducting and Coding Elite Interviews' (pp. 673–6); L. Woliver, 'Ethical Dilemmas in Personal Interviewing' (pp. 677–8); J. M. Berry, 'Validity and Reliability Issues in Elite Interviewing' (pp. 679–82); S. Werning Rivera, 'Interviewing Political Elites: Lessons from Russia' (pp. 683–8).
This brings together a set of presentations given as part of a short course on interview methods at the American Political Science Association, which give concise and helpful information about a variety of issues relating to interviews.

Weiss, R. (1994), *Learning from Strangers: The Art and Method of Qualitative Interview Studies* (New York: Free Press).
Offers a step-by-step introduction to the method of qualitative interviewing: sample selection, development of an interview guide, the conduct of the interview, data analysis, and preparation of the data. Includes examples of successful and less successful interviews.

Software

Friese, S. (2004), *Software Overview.* Available online as a PDF at:http://www.quarc.de/software_overview_table.pdf
A useful table comparing six software packages.

Matheson, J. L. (2007), 'The Voice Transcription Technique: Use of Voice Recognition Software to Transcribe Digital Interview Data in Qualitative Research', *The Qualitative Report* 12(4): 547–60. Retrieved from http://www.nova.edu/ssss/QR/QR12-4/matheson.pdf

Rettie, R. (2005), 'Exploiting Freely Available Software for Social Research', *Social Research Update* 48 (Winter). Retrieved 4 September 2006, from http://sru.soc.surrey.ac.uk/SRU48.html

References

Aberbach, J. D. and B. A. Rockman (2002), 'Conducting and Coding Elite Interviews', *PS: Political Science and Politics* 35(4) (December): 673–6.

Benoit, K. and M. Laver (2006), *Party Policy in Modern Democracies* (London: Routledge).

Bryman, A. (2015), *Social Research Methods*, 5th edition (Oxford: Oxford University Press).

Castles, F. and P. Mair (1984), 'Left–Right Political Scales: Some Expert Judgements', *European Journal of Political Research* 12: 73–88.

Dorussen, H., H. Lenz, and S. Blavoukos (2005), 'Assessing the Reliability and Validity of Expert Interviews', *European Union Politics* 6(3): 315–37.

Folkestad, B. (2008), 'Analysing Interview Data: Possibilities and Challenges', Eurosphere Online Working Paper Series, Working Paper No. 13 (December); available at: http://eurospheres.org/files/2010/08/Eurosphere_Working_Paper_13_Folkestad.pdf

George, A. L. and A. Bennett (2005), *Case Studies and Theory Development in the Social Sciences* (Cambridge, MA: MIT Press).

Glaser, B. G. and A. Strauss (1967), *The Discovery of Grounded Theory* (Chicago: Aldine).

Hooghe, L., R. Bakker, A. Brigevich, C. de Vries, E. Edwards, G. Marks, J. Rovny, and M. Steenbergen (2009), 'Reliability and Validity of the 2002 and 2006 Chapel Hill Expert Surveys on Party Positioning', http://web.me.com/rovny/Site/Academic_files/HoogheEtAl_EJPR_july2009.pdf

Jarvis, L. and M. Lister (2013) 'Disconnected Citizenship? The Impacts of Antiterrorism Policy on Citizenship in the UK', *Political Studies* 61(3): 656–75.

Kopecky, P. and C. Mudde (2002), 'The Two Sides of Euroscepticism: Party Position on European Integration in East Central Europe', *European Union Politics* 3: 297–326.

Miles, M. B. and A. M. Huberman (1994), *Qualitative Data Analysis*, 2nd edition (Newbury Park, CA: Sage Publications).

Pierce, R. (2008), *Research Methods in Politics: A Practical Guide* (London: Sage Publications).

Rivera, S. W., P. M. Kozyreva, and E. G. Sarovskii (2002), 'Interviewing Political Elites: Lessons from Russia', *PS: Political Science and Politics* 35(4) (December): 683–8.

Tansey, O. (2007), 'Process Tracing and Elite Interviewing: A Case for Non-probability Sampling', *PS: Political Science and Politics* 40(4) (October): 765–72.

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

Endnote

1. There are various types of computer-assisted interviewing software. Computer-assisted telephone interviewing (CATI) is usually employed in conjunction with random-digit dialling (RDD) as a means of approximating random sampling. Computer-assisted personal interviewing (CAPI) allows the interviewer to enter responses directly into an uploadable database using a hand-held computer. Audio-CASI uses voice technology to prompt the respondent with items and receive responses by voice.

13

Ethnography and Participant Observation

Chapter Summary

This chapter examines the principles of ethnography and participant observation: what they are, how (if) they became standardized as a research method, what form of evidence they constitute, and what place they occupy in the study of Politics. We look at the strengths of ethnographic fieldwork, and we try to identify what kind of material it produces and what aspects of social life it can reveal. We shall also discuss its weaknesses, especially issues of subjectivity, reliability, and generalizability. Topics discussed are:

- ethnography;
- participant observation;
- sampling;
- access;
- key informants;
- recording observations.

Introduction

Although participant observation was first developed by social anthropologists and sociologists, in recent years it has become a much more widely used research tool across the social sciences, and attention to political ethnographies in particular has grown (see Auyero 2006; Schatz 2009; MacKay and Levin 2015). Political ethnographies have been carried out in a wide variety of different contexts to study a range of topics, such as political parties and parliamentarians (Fenno 1978; Searing 1994; Rhodes 2011; Boswell et al. 2019), the judiciary (Latour 2010), local elites (Dahl 1961), international organizations (Weaver 2008), social movements and anti-politics (Blee and Currier 2006) and terrorist groups, drugs cartels, and the mafia (Parnell and Kane 2003; Michelutti et al. 2019). Political ethnography is also becoming more widely used in the study of International Relations (see Schatz 2009: 308; Lai and Roccu 2019). Gillespie led a team of ethnographers studying how British citizens, and British Muslims in particular, experienced securities and insecurities during the War

on Terror (see Gillespie, Gow, and Hoskins 2007; Gillespie and O'Loughlin 2009). Political ethnographies have also been carried out by Barnett (2006) on the US mission to the UN; by Pouliot (2007) on diplomacy and security; and Hopf (2002) on Russian foreign policy.

Political ethnographies have also been widely used in comparative research and have examined the political attitudes and behaviour of ordinary people living in different parts of the world, for example in India (Michelutti 2008), Senegal (Schaffer 1998), Uganda (Karlstrom 1996), and Chile (Paley 2001). Political ethnographies can be used to shed light on how marginalized social groups interact with modernization, globalization, and democracy, and how governmental policies and development programmes come to be reinterpreted, and accepted (or not) on the ground. Autesserre (2014) carried out an ethnographic study on conflict resolution and how international interventions work by following and working with international organizations in various post-conflict countries. Anderson (2014) carried out research with migrants, and acutally followed the migration route with them, rather than being immersed in a single place. And Adams (2013) provides an ethnographic account of reconstruction and recovery in New Orleans, showing how private companies exploited the disaster for profit.

Although many of these issues can be, and frequently are, studied using other methods, there is no substitute for getting out of the armchair, getting your hands dirty, and observing first-hand what it is that you are writing about in its natural setting. Indeed, one of the key strengths of participant observation is that it provides the researcher with first-hand experience of the subject they are writing about. As Richard Fenno (1990: 56) wrote, 'as long as political scientists continue to study politicians, some of us certainly will want to collect data through repeated interaction with these politicians in their natural habitats'. This chapter examines the principles of the ethnographic method, its application to political research, and the issues that confront researchers who use this approach.

Ethnography is often described as an approach rather than a specific method. Indeed, ethnographers can and frequently do employ a variety of methods, from **participant observation**, to archival analysis, interviews, and surveys. The term 'ethnography' comes from Greek and literally means writing about people, where *ethnos* means people or folk and *graphē* means writing. The term therefore encompasses and recognizes two important components of the approach. The first is to do with the study of people and how data are collected, and the second refers to how this data is then recorded and analysed (or written up). In this chapter we focus on both these components. First, we examine the method of participant observation, which is the method of data collection most closely associated with the ethnographic approach. Second, we examine how the observations and data from the fieldwork are recorded and reported.

The principles of ethnography and participant observation

Ethnography and participant observation can be defined in a number of ways, and these definitions frequently overlap with each other, and the two terms are frequently used interchangeably. According to Delamont (2004: 218)

> participant observation, ethnography and fieldwork are all used interchangeably . . . they can all mean spending long periods watching people, coupled with talking to them about what they are doing, thinking and saying, designed to see how they understand their world.

With reference specifically to ethnography, Brewer (2000: 6) writes that:

> ethnography is the study of people in naturally occurring settings or 'field' by methods of data collection which capture their social meanings and ordinary activities, involving the researcher participating directly in the setting, if not also the activities, in order to collect data in a systematic manner.

Likewise Hammersley and Atkinson (2007: 3) describe the ethnographic approach in terms of the:

> ethnographer participating, overtly or covertly in people's daily lives for an extended period of time, watching what happens, listening to what is said, asking questions—in fact, collecting whatever data are available to throw light on the issues that are the focus of research.

Both these definitions of ethnography include reference to multiple forms of data collection. To some extent then ethnography refers to a multi-method approach. It encompasses a plurality of methods, all of which are used in conjunction with each other to generate a deep, first-hand understanding of what people do in their social environment.

The main method that is employed to do this is participant observation. In order to observe what people do in practice, participant observation involves

> research based on the close-up, on-the-ground observation of people and institutions in real time and space, in which the investigator embeds herself near (or within) the phenomenon so as to detect how and why agents on the scene act, think and feel the way they do.
>
> (Wacquant 2003: 5)

The distinctive feature of participant observation, and one of the great strengths of the approach, is that data collection is carried out in real time. This means that the ethnographer has a direct, first-hand opportunity to observe what people actually do, what they actually say to each other, and how they actually interact with different institutions or political processes, rather than just relying on what people say that they do, which is not always the same as what they actually do (as we discussed with reference to survey measurement error in Chapter 11).

Participant observation (and ethnography more generally) therefore has a number of characteristics that overlap with other methods we have considered in this book (such as surveys, focus groups, and interviews) and a number of characteristics that are distinctive, particularly with respect to the role of observation. Whereas surveys (see Chapter 11) are based on the ancient art of asking questions to find out what people think, say, and do, participant observation is based on something rather different. It recognizes that what people say they do and what they actually do, can be, and frequently are, different. Accordingly, to get a 'true' sense of what people think and say and do, it is not enough to merely ask people questions and record their answers; it is also necessary to observe what people do in practice.

Participant observation, as the term implies, therefore involves the researcher living, and interacting, and participating in the daily life of the people (often referred to as **informants**) who she is studying. This can involve long periods of research, of many

months, if not years, in the **field** (the research setting). This prolonged period of time spent in the field allows the researcher to embed themselves in the setting which they are studying and really get to know their informants, understand their behaviour, and get detailed information about what they think and say and do. As Ingold (2008) puts it, this involves the researcher getting out of the armchair, and actually interacting with the people about whom they is writing.

This research on the 'everyday practices' of human behaviour has a number of key strengths. It is particularly useful for studying sensitive topics, where people may not directly reveal their real thoughts or opinions to a stranger. One such sensitive topic is witchcraft, which, although an important part of social life in many parts of the world, is generally not something that people feel comfortable talking about to strangers. Adam Ashforth's (2005) book *Witchcraft, Violence, and Democracy in South Africa* draws on three years of fieldwork in Soweto (South West Township), an apartheid-built black suburb of Johannesburg. In order to do his fieldwork, Ashforth learnt Zulu, Sotho, Xhosa, and the special brands of English Sowetans speak; he also learned to play the violin Zulu style, and was ready to defend his adopted brothers and sisters from recurrent threats of attack. Through first-hand observation, personal intervention, and long discussions with informants, Ashforth was able to build up a detailed description of life under the shadow of extreme violence. This helped him to arrive at two revealing conclusions: first, that no one can make sense of local South African politics without understanding the enormous part played by fears about, accusations of, and reactions to witchcraft in Soweto's (and, by extension, South Africa's) everyday politics; second, that no one can hope to deal with South Africa's devastating AIDS epidemic or build local-level democracy without confronting witchcraft directly (see Tilly 2006 for an extended discussion of these findings).

Participant observation is also well-suited to unpicking difficult-to-define or multifaceted political phenomena, where other research instruments such as surveys, interviews, or focus groups may provide too blunt an instrument to fully capture the diversity and meaning of a concept. For example, participant observation can shed light on and capture how Indians (Michelutti 2008), Senegalese (Schaffer 1998), Ugandans (Karlstrom 1996), or members of Chilean social movements (Paley 2001) understand what democracy means.

But perhaps the core strength of the method is to do with the act of observation. Participant observation allows the researcher to observe what people do on a day-by-day basis, and how their behaviour changes in response to different stimuli or different events. For example, using ethnographic, in-depth interviews, and document data on new and emerging social movement groups (SMGs) in Pittsburgh for twenty months before and after the 2004 US presidential election, Blee and Currier (2006) examine how SMG members think about elections and how groups decide to respond to national electoral campaigns. Participant observation also allows the researcher to observe first-hand how and why unplanned or spontaneous events emerge—such as a riot or a fight, a strike that develops in a pitched battle, or a simple quarrel that escalates and gets out of hand. This kind of research is often described as process tracing. It involves looking for evidence of the pressures, incentives, motivations, and decision-making calculus in any given instance of action (George and Bennett 2005; Parsons 2010). Moreover, by observing people over a prolonged period of time, it is possible to build up a much more reliable impression of what they really think than would be possible after just a brief meeting or interview.

Theory and hypothesis testing

A key purpose of participant observation is simply to describe what people think, say, and do. By spending a long time in the field, repeatedly talking to informants over time and in different social settings, and observing what people do, the method can provide a **thick description** of the social and political lives of the informants. In doing so, it aims to provide a valid and reliable representation of how the subjects under study behave and think politically. In some sense, then, the method is concerned with understanding, with meaning, and with interpretation. This sort of data is valuable in its own right. Colonial 'ethnographies' of different communities in India provide a rich and detailed description of social life in the nineteenth century, which would otherwise have been lost. Ethnographies constitute a rich source of historical data, and in addition can give voice to people who otherwise would not be heard, and shine a light into corners of the political world that would otherwise go undetected.

But at the same time, this ethnographic description and focus on rich textual detail is often thought to create a trade-off in terms of how far findings can be generalized to other contexts and to shed light on theory more generally. Ethnography and participant observation have therefore often been derided as only capable of producing 'thick description'. According to Hammersley and Atkinson (2007), ethnographic research comprises a strong emphasis on exploring the nature of social phenomena, rather than setting out to test hypotheses about them. Since ethnography tends to focus on the specific rather than the general, it is therefore often regarded as an inappropriate method to use to test general hypotheses.

Whether or not ethnography can be used to test theory is a debate which has rumbled on for many years. Indeed, as far back as the 1960s, Edmund Leach complained about the tendency of ethnographers to focus on the specific rather than the general, and he lamented, 'most of my colleagues are giving up in the attempt to make comparative generalizations; instead they have begun to write impeccably detailed historical ethnographies of particular peoples' (Leach 1961: 1). To be sure, ethnography can be used in a purely descriptive, inductive way (and often is), but equally it can be, and sometimes is, also used more deductively.

As we discuss in Chapter 2, it is easy to overstate the differences between inductive and deductive approaches. But, in general, ethnographic research begins with some problem or set of issues. At the very least, it starts from what Malinowski (1922) referred to as **foreshadowed problems**. Sometimes this is based on well-developed theory, sometimes it will be more exploratory, and sometimes it will lead to the formulation of specific hypotheses that are to be tested, or at least investigated.

Participant observation can be used for each of these purposes. That it is frequently not says more about the people who use the method than it does about the method itself. Thus, participant observation is concerned with producing descriptions and explanations of specific political phenomena, developing theories which may apply more generally, and applying theory to different social settings. It is a useful tool for generating theory in areas of research for which little is known, or where previous theory sheds little light. It is also a useful tool for applying theory, which is perhaps one of its most frequent uses. Ethnographers draw on insights from previous research and attempt to apply them to new contexts.

Ethnographic studies frequently take well-developed ideas (or theories) about hierarchy, or power, or division of labour, or social stratification, to name but a few, and see whether these ideas or theories apply in different contexts. If the theory is found to apply, and to provide a good framework for analysis, then the study in some sense corroborates or builds upon and develops the existing theory. But if the theory does not provide a convincing framework, then the study can be used to consider and reflect upon why it does not, and perhaps to propose some limitations to the original theory.

For example, Eliasoph's (1997) ethnography of different public and private spaces in a town on the US Pacific coast is a fantastic explanation of how political apathy is produced by contextual factors, and built up slowly through different phases and contexts of people's lives. She tests Noelle-Neuman's 'spiral of silence' theory, an explanation for why individuals feel unable to express their political views publicly, by spending years in this town getting to know groups and individuals, being a participant observer in social clubs, cake bakes, and town hall meetings, as well as talking to people in private.

The process of applying theory may, over time, with repetition, lead (indirectly) to the testing of theories, in that some theories may be found not to apply so many times that they lose their relevance and can essentially be falsified, while other theories are found to apply to many different contexts and so are given more empirical support. But the key point is that the testing of theory comes from the gradual accumulation of knowledge, rather than from any particular one-off ethnography.

Participant observation methodology

The main research instrument for data collection in participant observation is the individual researcher who actually undertakes the exercise. There is thus an inherently personal component to the research process. Better researchers make for better data collectors and data analysers. As Bernard (2006: 344) observes, participant observation is a craft, and as with all crafts, becoming a skilled artisan at participant observation takes practice. It also takes talent and specific personal skills. Participant observation involves getting informants to open up, being allowed into their lives, and having the opportunity to observe and record personal and private events first-hand. The extent to which the ethnographer is able to do this therefore depends in large part upon the ethnographer's personal characteristics, such as whether he is friendly and approachable, able to inspire trust and confidence; or whether informants think he is arrogant and standoffish.

Above all else, fieldwork involves the toil and dedication of the ethnographer, who has to spend thousands of hours in the field. Fieldwork relies on the tenacity and perseverance of the ethnographer, and the patience to sit through long, boring meetings in the hope that something interesting might come up. Ethnographers can spend months feeling as if nothing is happening and no progress is being made, only for things to suddenly take off. It can often feel like fieldwork is 99 per cent perspiration for that 1 per cent inspiration that unlocks the door to some specific issue of interest. Doing fieldwork is a long and arduous business, often in a very difficult environment, often among strange and unfamiliar people. It can be lonely and alienating, and is the most personally involved and challenging method of data collection that there is. Put simply, doing fieldwork is tough.

For many of these reasons, the idea of specific methodological training in participant observation was, for a long time, not taken terribly seriously. As Whiting recalls being told by a Yale University professor in the 1930s, method 'was a subject to be discussed casually at breakfast' (Whiting 1982: 156), not something worthy of a seminar. The skills for doing fieldwork were regarded as mysterious and unteachable (Bernard 2006: 343). Since so much depends upon the personal characteristics of the ethnographer, it is not really possible to know how someone is going to react to the pressures of doing fieldwork until they are out in the field, actually doing it. Ethnographers were therefore expected to basically just get on with it, and learn how to do participant observation on the job.

Although much of this still holds, it does perhaps represent something of a romanticization of what it means to do fieldwork. Participant observation is both a humanistic method and a scientific one (Bernard 2006: 342), and as such it entails a number of methodological principles that can affect the quality, reliability, and validity of the data that is collected. In recent years, there has been a growing recognition of this, and there are now many textbooks specifically on ethnographic methods (see Hammersley and Atkinson 2007 for a particularly in-depth example), and many undergraduate and graduate courses that address different methodological and ethical considerations connected to the process of collecting and analysing data.

Participant observation involves the researcher making a number of methodological decisions, and the decisions that are made can influence the quality of the data that is collected. These decisions face the ethnographer every day, and whereas it is not possible to discuss all of them in detail, it is nonetheless important to highlight some of the main issues that need to be confronted, and discuss their implications for the quality of data that is generated. Broadly speaking, participant observation involves three interrelated steps. The first step refers to case selection and gaining **access** to the field site. Participant observation relies (usually) on the detailed analysis of a single case study. The criteria and justification for the choice of case (and type of case) that is selected is very important, since the type of cases you choose to analyse can influence the answers you get to a particular research question (see Chapter 9). This problem is even more of an issue when only one case is being examined. The second step refers to issues related to carrying out research in the field and collecting data. The role the investigator adopts, the contacts they make, the informants they observe and speak to, and the ways in which this is done, all have a substantial bearing on the quality, reliability, and validity of the data that is generated. Finally, the third step refers to recording observations, and how the data that is collected is written up, coded, and used for analysis.

In some sense, these different issues are analogous to the sampling and measurement issues that we discussed with reference to survey research in Chapter 11, and to a greater or lesser extent the same issues underpin all empirical research, whether it is qualitative or quantitative. We have to be aware about how we select our cases, how we collect our data and code it. The decisions we make at each step can have a strong bearing on what our final results look like. Participant observation may prioritize internal validity over external validity, but it certainly does not ignore the latter. But even to be internally valid, it is necessary to pay special attention to how the data is collected and coded within the domain

of study. It would be problematic, to say the least, if another ethnographer carried out research on exactly the same topic in the same location at the same time and came to very different conclusions.

Choosing a research site

In a sense, all ethnographies are a form of case study (though the reverse is not true). And, as with any case study, the selection of the case (or field site) poses a number of challenges, and there are many issues that need to be considered. Hammersley and Atkinson (2007: 28) suggest that case selection can proceed in one of two general ways: (1) the issue, or research question, or foreshadowed problem comes first, and a case or setting is then selected in order to investigate the topic; or (2) in some special circumstances, the setting itself may come first. A spontaneous opportunity may arise to investigate an interesting situation or group of people, and the research questions spring from the nature of that setting. For example, Pieke (1995) recounts how five months into his fieldwork in Beijing, he witnessed the emergence of the 1989 Chinese People's Movement. He thought this presented an interesting opportunity, and so started to carry out research. The setting thus came first, and his research questions were thus largely driven by his choice of case, rather than the other way round. However, instances of this type of 'opportunistic research' (Riemer 1977) are relatively rare, and more usually the choice of research setting (at least initially) is guided by the research question and the issues or hypotheses that the ethnographer wants to investigate, though, of course, these may change somewhat during the course of fieldwork.

One of the key problems that ethnographers face is that they sometimes find that the field site they have selected is not an appropriate arena to investigate what they had initially hoped to explore. To a greater or lesser extent this problem is relatively frequent, and it is widely accepted that the collection of primary data often plays an important role in the development of the research question. What an ethnographer initially sets out to investigate may not therefore always be exactly the same as what he ends up investigating. Although this isn't always a problem, and indeed it can sometimes lead to the elaboration of new and exciting research questions, if the initial question or issue was thought to be worth investigating in the first place, then it is worth seriously thinking about how this can be done.

In particular, there are two interrelated issues that need to be addressed. The first is to establish the type of site that the researcher plans to study, and second, once this is done, to decide which site in particular to select for investigation. That is, before selecting the site, it is first necessary to think about the population of potential sites that would be eligible (or appropriate) cases for selection. The more thought that is put into this at the planning stage, the less risk there is of ending up in a place that is ill-suited to examining the desired issues.

Types of research site

Early ethnographic studies using participant observation tended to be the preserve of anthropologists, who often worked in small, isolated, self-contained communities that had little to do with the outside world and state structures. However, ethnographers have

increasingly turned their attention beyond village settings to investigate social and political phenomena in a wide variety of settings, from urban ethnographies to ethnographies of political parties, institutions, and social movements. Indeed, some of the most in-depth ethnographies rely on fieldwork in multiple settings. For example, Michelutti's (2008) study on how democratic ideas and practices were reinterpreted among the Yadavs, a historically marginalized community in India, was carried out in a neighbourhood where many Yadavs lived, in the temples and religious festivals that they attended, in the political parties to which they were affiliated, and in the cultural associations of which they were members. Although the main research setting was the urban area where they lived, the other settings stemmed from this, linking the local area to the wider region, the state, and even to the nation. In this sense, ethnography is often described as holistic, in that it examines how what people say and do plays out in different settings and domains over time. Linked to this, recent discussions of multi-sited approaches seek in a similar way to link the local to the regional and even the global (Marcus 1995; Comaroff and Comaroff 2003). Multi-sited approaches carry out fieldwork in different research sites (Coleman 2011). For example, research on the Punjabi diaspora might carry out research in the Punjab, on people living there, and then carry out research in a different site in Southall, in the UK, where large numbers of Punjabis live.

By contrast, other studies, particularly in political research, have tended to have a more specific focus and are often primarily based on fieldwork in organizations and institutions, such as political parties, the civil service, parliament, the EU, and non-govenmental organizations (NGOs), or loose networks, such as gangs, religious extremists, or freedom fighters. Members and participants may not all live in the same place, and so are primarily linked through the organizations. For these cases, the main field site is therefore the organization, such as the local political party headquarters, and other research sites may then stem from this (see, e.g., Bale and Dann's 2002 study of New Zealand's Green Party). Increasingly, scholars also carry out digital ethnography on the internet to study online communities (Costello et al. 2017). See Box 13.1. The definition of the primary research site and justification for why one type of site is appropriate rather than another is therefore a crucial first step, which needs to be related explicitly to the research question and objectives.

Selection of research site

Having defined the type of research site that is appropriate for the research project, the next step is to decide which specific site is to be selected for the study. Often practical considerations form a prominent basis for selection, and although these are important, they should not be entirely divorced from theoretical and methodological considerations. The factors that make access to a site easier might also make it peculiar in some important respects, which means that it might not be very representative or typical of the population of interest. For example, although there is a great deal of academic interest in the shanty towns, or *barrios*, that surround major cities in Latin America, these locations are often dangerous and violent places which make it risky for a researcher to just turn up. There is a tendency among researchers, therefore, to focus on the *barrios* which are relatively safe, and where there are active NGOs and religious organizations that can help with access arrangements and introductions. Although this paves the way for some incredibly valuable

BOX 13.1 Digital Ethnography

Digital ethnography—or netnography—is increasingly used by scholars to study online communities and social media spaces. The term netnography was first coined in the 1990s by Kozinets (1998) to refer to the qualitative study of online communities, which were beginning to emerge at that time around discussion groups or chat rooms. As with more traditional ethnography, netnography involves participant observation. As an approach it is particularly well-suited to studying sensitive topics or even illegal activities, where the anonymity of the online environment may facilitate more frank and open discussion. Kozinets (2015: 88) describes it as having a 'voyeuristic quality' because the approach can be used to study situations or topics which might otherwise be more difficult to study face-to-face. Examples include studies on people who have had cosmetic surgery (Langer and Beckman 2005), 'fat activitists' (Gurrieri and Cherrier 2013), or people with specific health issues or problems (Bratucu et al. 2014). As the study of online communities and social media spaces has expanded, the term netnography has been used (and abused) in a variety of different ways, with some scholars simply choosing to focus on data that is easy to collect and analyse, minimizing their own engagement with the members of the online communities they are studying (Costello et al. 2017). However, at its core, the standards and practices that make for a good netnography are very similar to those that make a good ethnography. According to Kurikko and Tuominen (2012: 13) 'some of the most important standards of quality in netnography are immersive depth, prolonged engagement, researcher identification, and persistent conversations.' However, carrying out research online like this poses some ethical issues. Members of online communities may not want to be studied. LeBesco (2004) reported that in a single month eight researchers tried to gain access to a particular online community site and all but one were rejected by the group. And Hudson and Bruckman (2004) report cases where people in chat-rooms reacted with hostility when they were aware of being studied by researchers. Kozinets (2002) outlines four key ethical guidelines for netnography: (1) The researcher should always disclose their identity; (2) the researcher should guarantee confidentiality and anonymity; (3) the researcher should seek and incorporate feedback; and (4) the researcher should ask permission to directly quote any specific posting. While these ethical guidelines are well-suited to pretty much any type of research on human subjects (see Chapter 6), Langer and Beckman (2005) argue that they are not applicable to 'passive' netnography which takes place in a public forum, and is more like a content analysis of text data. In addition to these ethical issues, there are also important legal issues which should be remembered when doing online research. Saying 'it was for research' or 'I was writing a book' is no defence for breaking the law and accessing illegal websites.

research, at the same time *barrios* which have these organized links with civil society are likely to be somewhat different from the ones where NGOs and religious organizations are not present. This is an important consideration when thinking about how far findings from the research may be generalized.

A useful starting point for guiding case selection should therefore be to try and define the population of eligible cases for selection. What are the theoretical criteria for selecting a particular case? And what other potential cases might fit these criteria? Is the case chosen because it is intended to be typical or because it is an outlier? For example, if the research project is on Hindu–Muslim relations in India, are you interested in examining a neighbourhood that is characterized by particularly high levels of conflict, or particularly low levels of conflict? Thinking about these questions will help to identify the population to which your findings might apply, and to set the parameters for potential generalizations that may stem from the research.

Access to research site

Having selected a site at which to conduct fieldwork, one of the most difficult steps in carrying out participant observation is gaining access to the research site. It is all very well to have a clear idea about the type of place where you want to do research, and even to have selected a potential site. But if it is not possible to gain access to the field site, then it is back to the drawing board. Access issues can create problems in two main ways. First, if access is not possible, it may introduce response bias similar in nature to the selection bias discussed earlier. Places where you cannot gain access may be different in important respects from the places where you can (or can more easily) gain access.

Second, access issues can also affect the way in which you carry out research. How you enter the field can shape how you are perceived by informants, and so your point of entry may have long-term consequences for your research project. One obstacle ethnographers frequently face is to overcome the suspicion of the people they are studying. For many informants, the purpose of ethnography is hard to fathom. Why would anyone want to study us? It is sometimes much easier to assume that the investigator is a spy rather than a political scientist. With time and hard work, these suspicions can be overcome, but the extent to which this is possible and the amount of time it takes may be influenced by the level of distrust in the first place. And on this score, first impressions can count for a lot. In this section we discuss some of the strategies that can be employed to gain access to the research site, and the impact that these strategies may have on the way in which research is conducted.

The issues surrounding access vary in a number of ways according to the object of investigation. This is often discussed in terms of whether the research setting is **open** or **closed** or a public or non-public setting (Hammersley and Atkinson 2007). The distinction between the two types of setting is relevant for gaining access because a closed setting may require the permission of someone in order to gain access. Access to closed (non-public) settings, such as formalized groups or institutions, political parties, law courts, the police, the European Union, United Nations, development organizations, NGOs, to name just a few, present a number of problems, since some sort of formal permission may be required in order to carry out research. Bryman (2015: 428) suggests a number of strategies for gaining access to closed (organizational) settings, such as using friends, contacts, or colleagues to try to get support from someone within the organization. In particular, it is highly likely that you will need permission to do research from a **gatekeeper**—senior management in the organization. Gatekeepers may be wary about letting researchers snoop about in the organization, particularly if they have concerns as to whether the organization will be presented in a good light or not. Similarly, government permissions, or permits, are often needed to carry out research in tribal areas or reservations, and these permits can be difficult to come by, particularly if the research is on a sensitive topic. It is often then a good idea to offer gatekeepers something in return, such as a report or summary of your main findings. In dealing with gatekeepers, Bryman also suggests that it is a good idea to provide a clear explanation of the aims and methods of the research, and the amount of time it will take (and take away from people doing their job or other duties). It is also important to negotiate. Full access may not be forthcoming, but with careful compromise it might be possible to have some sort of limited access that nonetheless enables

the researcher to investigate their topic. As Schatz (2009: 307) writes, access is a 'sliding scale' on which the political ethnographer strives to 'achieve the *nearest possible vantage point* to study a given problem'.

Where permission is needed, it is obviously important to get the permission giver—or gatekeeper—on board. However, it is also important to bear in mind how the compromises you make (if any) and the relationship you develop with the gatekeeper will affect your research, and how it will shape how you are perceived by informants once you have gained access. The extent to which this is an issue may depend upon the type of setting, and the hierarchy within the organization, and how the gatekeeper is viewed by those with whom you will be doing research. For example, if you are working in a factory, and have been given permission to do research by an unpopular member of senior management, then there is the risk that by association you will be viewed by the workers in an unfavourable light, and perhaps seen as a spy.

Access to open settings (public sites) is in theory more straightforward, since there aren't the same bureaucratic obstacles that need to be negotiated. But in practice it is beset by many of the same issues that face those seeking to gain access to private settings, particularly when the object of study is defined as a hard-to-reach population. If the site is a neighbourhood, or village, or other public space, then it is not necessary to obtain formal permission to be there, and so the researcher can arrive, check out the location, and simply start observing what is going on. To facilitate this process, researchers often choose field sites where they have some prior personal link or contact.

However, some types of open public sites can present substantial access problems, particularly if the group in question is hard to reach, or involved on the periphery of illegal activity, or indeed completely immersed in it, such as mafia, terrorist groups, or freedom fighters. For these inaccessible groups, it is often necessary to develop contacts who can provide an introduction to the group in question. However, obtaining these contacts is far from straightforward. They rely on trust and so are often only made after many years of hard work.

The role of the ethnographer

One obvious way to reduce some of these access problems is to do away with the need for permission in the first place, and for the investigator to assume a **covert** role, in which they do not reveal the true purpose of their research, or even that they are a researcher at all. This strategy means that the researcher does not face the problems of having to persuade people to give their permission to be studied, and the researcher doesn't have to face awkward questions about why they want to do the research either. This distinction between **overt research** and covert observation also influences the way in which data is collected, and the role that the ethnographer adopts in the field. Broadly speaking, there are three main types of role the researcher can adopt: the participant, the participant observer, and the observer.

The participant is a fully functioning member of the social setting and their identity is not known to the informants. The participant is a covert observer. It is widely documented that the act of being studied can alter the behaviour of the subjects under investigation.

BOX 13.2 Ethical Issues in Participant Observation

Ethical issues in participant observation and ethnography are extremely important. These issues have received a great deal of attention, particularly in political anthropology. All of the ethical issues discussed in Chapter 6 are relevant, but the issue of voluntary participation and informed consent are perhaps the most contentious. Participant observation is often carried out in a range of different settings, both public and private, and fieldwork may be carried out by means of casual conversations with a wide range of informants, some of whom may be well-known to the researcher and seen regularly, others of whom may only be met during fleeting encounters. In these contexts gaining informed written consent can pose something of an obstacle to doing research. Requiring signatures on printed forms may damage rapport with informants, and even make informants suspicious. More importantly, written informed consent may itself pose ethical issues when informants are asked about their involvement in illicit activities or about sensitive issues, and are anxious about protecting their anonymity. Gaining consent is therefore often understood as an ongoing process, which is verbally negotiated. It is important that the researcher is as open and transparent as possible about their identity as a researcher and explains the objectives of their work. Gaining consent should not be regarded as a one-off event and should be re-established over time during the course of fieldwork. Another thorny issue in ethnographic work is the issue of privacy. At one level, in ethnographic research, it is easy to use pseudonyms for the informants so that they cannot be identified by name. But this may not always be enough to ensure privacy. Informants may be identifiable by information you have revealed about them, particularly if it is tied to a specific field site, such as a town or organization. Some ethnographers therefore prefer to make the field site anonymous. For example, they may invent a fictitious name for the town they are working in. But people in the town may still be able to recognize and identify informants, particularly if it they are prominent locally. And it may still be possible to identify the town from particular characteristics relevant to the study (e.g. if it witnessed a major episode of ethnic violence; or is the site of a major religious shrine). As ethical guidelines become tighter and more strictly enforced, there is a danger that ethnography could become legislated out of existence.

This is known as reactivity. One way to reduce this reactivity is for the researcher to adopt a covert role so that the informants don't know they are being studied. The behaviour of informants will therefore be more natural, and this can often lead the researcher to gain much closer access. For example, Humphreys (1970) pretended to be gay in order to enter the world of those he was studying.

The participant observer is the most common approach used by the investigator in ethnographic research. It is much the same as the participant approach, in that the researcher participates directly in the informants' lives, but differs in that it is overt, and the informants are aware of the researcher's identity. By contrast, the observer (or observer-participant) is focused on observation rather than participation. The researcher acts mainly as an interviewer, or carries out unobtrusive observation.

Although there are some obvious attractions to the covert approach, particularly when the researcher is concerned that permission might not be given if asked for, or when the researcher is concerned that knowledge about their identity might alter the way in which informants relate to their, there is also a number of ethical issues related to carrying out covert research to do with deception and informed consent (see Box 13.2). It is therefore an approach that is not widely used anymore.

Doing participant observation

Once in the field, there is a number of issues that the researcher faces in terms of how they can do participant observation, and how this will influence the quality of data that is generated. In participant observation, the main instrument of data collection is the individual ethnographer. The reliability and validity of the data that is collected therefore depend in large part upon how they behave, who they talk to, what they observe, and where they go. In addition, there are also all sorts of personal issues—such as the gender, ethnicity, and age of the researcher—that may influence how they are perceived by their informants, and shape their role as a researcher within the community. This can introduce considerable subjectivity into the research process, which may bias the conclusions that are reached (see Chapter 3 for an extended discussion of this).

Having gained access to the field site, a requisite of participant observation is to build and develop a relationship with the people who are being studied. Part of this comes simply with time and contact. One of the most important factors that influences the quality and reliability of the data generated is therefore related to the amount of time spent in the field. All other things being equal, the ethnographer is likely to have much more reliable data if they have been doing fieldwork for twelve months than if they have only been doing fieldwork for twelve days. It is only by observing events and talking to people over a relatively long period of time that the ethnographer can build up a picture of how the community under study really thinks and behaves.

But the amount of fieldwork that is needed can also vary substantially according to the researcher's familiarity with the method of participant observation, the topic of study, and the location of fieldwork and the people. Since many of the skills of doing participant observation are learned on the job, the first major piece of fieldwork by an ethnographer might be as long as eighteen to twenty-four months, but subsequent fieldwork might be much shorter. With practice, the ethnographer will be able to collect data more efficiently, they will have a better eye for what to look out for, and possess more finely tuned research skills. Moreover, if the ethnographer is going back to the same place as a previous fieldwork, then much of the background data is already in the bank. They already know the people and the language, and so can collect good data in just a matter of days or weeks.

In addition, in order to develop a relationship with informants, it is necessary for the ethnographer to learn the language of the people they are studying. This is a core skill in its own right, and the linguistic competence of ethnographers is a key part of being able to do research effectively. Although some ethnographers use interpreters, to pick up sensitive data it is advantageous to know the local language. 'Fluency in the local language doesn't just improve your rapport; it increases the probability that people will tell you sensitive things . . . and that even if people try to put one over on you, you'll know about it' (Bernard 2006: 361).

Choosing informants

One of the great strengths of participation is that it is thought to have high levels of 'internal validity'. This means that the researcher can be confident that the findings from the research are valid with reference to the domain of study (field site), even if they are less confident

that the findings can be generalized more widely to other potential sites. However, this point can easily be overstated, and should not be taken for granted. There are a number of issues that influence the internal validity of ethnographic research, and, without careful attention, it is easy to produce data that are neither internally nor externally valid.

As with surveys, doing participant observation involves sampling. Even just hanging around and chatting to people is a kind of convenience or volunteer sample (see Chapter 11). Ideally, the ethnographer would speak to everyone in the population of interest (field site), and observe everyone equally. However, in practice this is rarely possible. The ethnographer might speak to some people more than others and spend time with some people more than others. The sample of informants that the ethnographer primarily relies upon may therefore only be a relatively small subsection of the entire population. This is particularly the case for urban ethnographies, where hundreds if not thousands of people may live together in relative proximity, or for ethnographies of relatively large organizations that might contain many members. In these situations, the ethnographer needs to think very carefully about whom they spend time with and why, as error and bias can easily be introduced into the sample, which will mean that it is not representative of the population of study, and so the internal validity of the findings will be compromised.

For example, suppose a researcher is conducting an ethnography of the House of Commons, and is doing participant observation among British MPs. Part of the research will involve watching what MPs do and say in public places, but part of the research will also be carried out in more private settings, where in-depth conversations and interviews with MPs can take place. Perhaps also the researcher will track different MPs, and get permission to observe who they have lunch with, who they talk to and meet, and so on to shed light on the links and contacts that MPs have with business and lobby groups and different social organizations. Basically, the ethnographer will try, as much as possible, to live the life of an MP in order to get a detailed first-hand impression of what they do. But undoubtedly the researcher will spend more time with some MPs than others. Perhaps they will only spend time with backbench MPs because senior ministers and shadow ministers may be less prepared to give full access to the researcher because they are concerned about leaks, confidentiality, secrecy, or are simply too busy. So already there is some selection bias, and what started off as a study of MPs is now becoming a study of backbench MPs. But, even among these backbench MPs, some might be more helpful and accommodating and more prepared to give up their time. There is always a temptation to be drawn towards the cooperative informants and to spend time talking to the people with whom you get on best, who are most helpful and approachable, and who make the job of the ethnographer easier.

Some informants may therefore be more useful or helpful than others, and certain informants—or **key informants**—may become more central to the collection of data than others. These key informants can be a great asset. They may understand the purpose and aims of the research, and be able to direct the ethnographer to situations, or places, or people who may be useful for research. But they also carry risks. Over-reliance on key informants can create a different set of problems, to do with the reliability and validity of the information that is generated. As Bryman (2015: 432) puts it, a reliance on key informants may mean that 'rather than seeing social reality through the eyes of members of the social setting, the researcher is seeing social reality through the eyes of the key informant'. It is

therefore very important for the ethnographer to continuously bear in mind the representativeness of their informants, and to actively seek out different voices and speak to those who may be less forthcoming, as well as those who are helpful and amenable. The consequence of failing to do so (or not doing so adequately) is to introduce error and bias, which can undermine the internal validity of the study.

Studying with informants

Because the nature of participant observation relies on so much social interaction, Ingold (2008) characterizes it as study *with* a particular group of people, rather than just a study *of* people. The social interaction is thus an important, but also potentially problematic, aspect of participant observation, as the researcher aims to become 'accepted' and 'trusted' by those that they are working with. As Whyte (1993: 303) argues, 'if people accept you—you can just hang around and you can learn the answers in the long run without even having to ask questions'. Building trust is thus a key part of the research process. It creates an environment where informants can speak frankly about sensitive and private issues without embarrassment, where they have the confidence to express their own views rather than the views they think they should hold or are culturally expected to hold, all of which allows the ethnographer to collect rich, detailed, valid data. In addition, the more an ethnographer is accepted by the community they are studying, the less likely informants are to respond to the presence of the ethnographer. Rather than sticking out like a sore thumb and constantly drawing attention to themselves, the ethnographer can slip into the background, become part of the mundane, and observe quietly what goes on around them. But if the ethnographer is not really trusted—or is viewed with suspicion—then there is a greater risk that the ethnographer, just by their very presence, can influence what happens in the research site. Informants will be guarded around the ethnographer, will change their behaviour when they think they are being observed, and act in an artificial way.

This is known as reactivity, which can be a major source of error in ethnographic research. Reactivity occurs when the very presence of the researcher alters the way in which informants act. There are two main strategies for dealing with this (see Bernard 2006: 425). The first is training. It is not possible to eliminate observer bias completely, but there is a wealth of evidence to suggest that training and experience make people better—more reliable and more accurate—observers (Hartmann and Wood 1990). The second is to build trust. When informants are comfortable in the presence of the ethnographer, they become more unguarded and less likely to change their behaviour when the ethnographer is around. This helps to reduce (though it never entirely eliminates) the problem of reactivity. Becoming accepted and gaining trust is thus an important part of doing fieldwork. In doing so, the ethnographer will have greater opportunity to observe and talk about things that are perhaps on the borders of social acceptance or legality—and enter worlds of criminality, corruption, usury. Informants will also be more likely to reveal personal information and allow the ethnographer to ask direct and personal questions, examine sensitive areas of research, and be invited to private settings.

But this acceptance can also be a double-edged sword. Becoming accepted within one part of the community of study can also isolate the researcher from other groups. For example, a researcher interested in studying Hindu–Muslim relations in India might find that

as they become accepted by Hindus, they become distrusted by Muslims. Their research therefore becomes skewed—or biased—towards one side of the story. Similarly, during Stacey et al.'s (1975) research on local party politics, the research team hired offices in the same building as the local Labour Party headquarters. This gave the researchers good access to Labour Party workers, and they soon built up good relations with them. But in the process of doing this, they realized that they were now being viewed with suspicion by Conservative Party workers, who thought that the researchers were too close to Labour, and somehow 'on their side'. These problems can compound the access problems discussed previously, if the group that the researcher is perceived to be closer to is also the group that they gained access through.

A related problem associated with building trust and acceptance is to do with the nature of the researcher's relationship with the people they are studying. Although the ethnographer needs to cultivate good relations in order to collect good data, they also need to be able to maintain some emotional distance from the people they are studying so they retain the capacity for critical analysis. As the researcher becomes more immersed in the community of study, they may find it harder to be **objective** (see Chapter 3). In extreme circumstances observers may even 'go native' and identify so much with the subjects that they lose the capacity for objective criticism. They may even become an advocate for the group, and campaign and lobby on their behalf.

It can be incredibly difficult to stay neutral and impartial when the act of data collection involves so much close social interaction in order to gain the trust of informants. This can make it difficult to present a 'warts and all' betrayal of real life, and the ethnographer may not want to present their informants in a bad light, particularly if they would like to go back to the field site to do more research in the same place. The practical difficulties of maintaining this balancing act can be hard to pull off, and there is no easy solution. While it is desirable to become accepted within the group, there can also be unintended consequences of doing this. Indeed, it can sometimes feel like an act of betrayal and raise ethical issues. The art of gaining acceptance is often described as a euphemism for 'impression management' or deception.

Identity of the researcher

The ability of the ethnographer to build a relationship with informants, and the type of relationship that they are able to build, depends not only on what the ethnographer does in the field, but also to a certain extent on who the ethnographer is. There is a large reflexive literature on how the identity of the ethnographer influences how informants perceive them. In particular, much of this discussion has focused on the sex of the ethnographer, but other social characteristics, such as age, ethnicity, nationality, and even marital status and whether or not the ethnographer has children, have all been identified as having an important impact on how the researcher is received by informants and the kinds of access to different social worlds that they are granted.

As far back as the 1930s anthropologists such as Margaret Mead had already started to highlight the impact that the gender of the ethnographer could have on the collection of data in terms of the consequences it had on (i) access to information; and (ii) on how others perceive the ethnographer. Some social characteristics have a fairly restrictive impact on

what the ethnographer can or cannot study. Men, for example, would not be able to carry out participant observation with Muslim women in *purdah*, since they would be prohibited from being in the same room as them. Similarly, much of the discussion on gender has also examined the obstacles that gender stereotypes create for women researchers, who have often been restricted to the study of domestic realms involving other women, children, and elderly people (Hammersley and Atkinson 2007). The gender of the researcher can thus both help and restrict access to different areas of social life, depending upon the cultural norms of the informants. What you can and can't do as a man or woman (or where you can and can't go) is more fixed in some cultures than in others, and in many cultures there is a certain amount of elasticity in gender roles that creates lots of individual variation.

The gender of the researcher can also have a less obvious impact, and influence the way in which informants respond to questions.

In addition, the ethnicity and nationality of the researcher can also influence how informants react. Sometimes it can actually be an advantage if the ethnographer belongs to a different ethnic or national group than the informants. For example, Hannerz (1969) carried out research in a black ghetto in the United States, and commented that although he was often jokingly referred to as 'blue-eyed blond devil', the fact that he was Swedish rather than American helped to distance him from other whites, who were not favourably received by his informants. Similarly, an Indian Brahman studying Indian Dalits, or an upper-class English man studying Welsh coalminers might take much more 'baggage' into the field with them than would someone from a different country less closely tied to the people they were studying.

The extent to which the ethnographer shares social characteristics with the informants can therefore be both a help and hindrance to building a relationship, depending upon the context. It is not therefore that the ethnographer should allow these factors to determine what they do in the field, but rather that they should be aware of how they might influence both what they can do and how people respond to them.

Critics of ethnographic research often challenge it as a method which is inherently subjective (see Chapter 3). Accordingly, these subjective influences can affect what type of data is collected, how it is collected, and from whom it is collected, all of which means that the data collected and the findings that come from them may tell us more (or more than we would ideally like) about the researcher who carried out the project than about the people or place that were actually being studied. This creates a serious problem in terms of **replicability**, which is one of the hallmarks of robust research (see Chapter 2). If different researchers studying the same phenomenon in the same place come to different conclusions, how can the findings from either be trusted? In what sense, if any, can the findings from ethnographic research therefore be regarded as internally valid?

Political ethnographers have long struggled to answer these questions (Schatz 2009; Wedeen 2010). For some the task becomes too daunting, and so they retreat ever further into self-reflexivity, producing 'descriptive analysis of the most limited, self-referential sort—explanation phobic' (Comaroff 2010: 526) and reducing 'ethnography to a solipsistic literary practice, one so obsessively reflexive as to be of no interest to anybody outside of itself' (2010: 525). For others, the only way to get around these problems is to incorporate ethnographic research within other methodologies, so for scholars like Laitin (2003, 2006) 'narrative' approaches (such as ethnography) are 'by themselves inadequate, but when combined

with large-N statistical work and formal models . . . can help generate robust findings' (Wedeen 2010: 259). See also discussion in Hopf (2006) and Pachirat (2009). Both these views represent somewhat extreme positions. And whereas it is probably true that subjectivity is inherent in ethnography, as it is to a greater or lesser extent in all social research, particularly qualitative research, this should not be overstated. As we discuss in Chapter 3, no approach or type of research ensures a completely value-free process of inquiry, and none can therefore free researchers from the need to be explicit and self-critical concerning their own underlying assumptions and values and behaviour, and how this might influence the research process. But this self-criticism is only of any use if it helps to guide the research process.

Recording observations

Ethnography involves two distinct but related processes: collecting data through participant observation, and recording and writing up that data. The quality of data that is collected in an ethnography rests not only on how the fieldwork was conducted, but also on how the fieldwork was recorded. Some forms of data can be relatively easy to record. For example, an interview or speech can be digitally recorded and transcribed (see Chapter 12). But other types of data are more difficult to record, and the distinction between recording or preserving data and analysing that data becomes blurred. Indeed, in many ways it is almost unavoidable that researchers will interpret their data—and to a certain extent construct their data—because as they write up and record what they have seen or heard, the process of recording data is filtered through the ethnographer. It is not therefore a simple record or representation of objective information, but a construction of that information that is in part interpretation, part analysis, and part observation. In this section we examine how different types of data can be recorded in preparation for analysis.

Field notes

Field notes are the traditional way of recording observational and interview data in ethnography. Originally they were handwritten in field journals, but now there is a wide variety of computer software programs that can be used to record and organize field notes. For a detailed discussion about the production of field notes, see Emerson, Fretz, and Shaw (1995) and Sanjek (1990). The writing of field notes is something that needs to be done with great care and attention. Poor quality note-taking can severely hamper the research process, and undermine the study. It is important that notes are comprehensive, but at the same time field notes are always selective. It is not possible to record everything, and so the ethnographer must prioritize what is interesting or relevant to the research project. As Hammersley and Atkinson (2007) note, taking good field notes is not just about writing down what you see and hear, but about knowing what to write down, how to write it down, and when to write it down.

The *when* is extremely important. As a general rule, the quality of notes diminishes with time, and the researcher should therefore aim to make notes as soon as possible after the observed action or conversation has taken place. Details can easily be forgotten or

confused, and the longer the passage of time between the event and the record of the event, the greater the likelihood that this will happen. Ideally, then, notes should be made in real time, as events or conversations are taking place. But this is not always practical, and can be disruptive to the activity of carrying out participant observation. It is often necessary to wait until a time when the researcher is not directly involved in carrying out the research, and can spend time writing up the day's activities and expanding upon any brief notes that have been jotted down during the course of the day. Bernard (2006: 387) recommends spending two to three hours per day doing this.

In some sense, field journals and recordings provide the raw data that the researcher then uses for analysis, though the distinction between data and analysis is not always clear-cut. Bernard (2006) distinguishes between descriptive notes and analytical notes, though in practice there is likely to be considerable overlap between the two. In preparation for writing up and analysing field notes, there are now many software programs that can be used to index and code the data. It is important to draw on the full breadth of the fieldwork data when writing up the results, rather than selectively using only the data and evidence that is convenient to the argument, and modern software programs can help to ensure that this is done more systematically.

One of the main problems with ethnographic research is that it is very difficult for the reader to know whether this has been done, which makes it difficult to assess the reliability and validity of the research. In recent years there has been some discussion of whether ethnographic data should be deposited in archives and made available, similar to the standard practice with survey and interview data. This would help to remove the veil of secrecy that surrounds much of fieldwork and ethnographic data collection. The advantage of this is that it would create a firmer distinction between data and analysis. People would be able to see how data was collected, what data was collected (and what was not) and how it was then analysed. This would create a stockpile of evidence and information that could be re-analysed—perhaps with differing conclusions from those of the original author. This could help make better use of data which are already collected. Doing this would no doubt improve the transparency of data collection, and reduce the probability of ethnographers 'making up' evidence to fit their argument.

For example, the 2004–7 Shifting Securities ethnography project (Gillespie et al. 2007), led by Marie Gillespie, which investigated ordinary people's understandings of security, was required by its funding body to deposit all ethnographic transcripts and reports on a secure website (http://www.mediatingsecurity.com). As a collaboration between eighteen ethnographers, the project website became a resource for the project team itself to exchange information, preliminary papers, and of course the notes and reports of others. This allowed for the coordination of research questions and emerging hypotheses across the team. It also enabled the team leader to pay attention to issues of internal validity within the study, since patterns and outliers could be observed as findings came in across the eighteen individual ethnographers' studies. Moreover, it allowed researchers working across the fields of political and social science to access this data, compare it with their own studies, and use that data to develop their own projects. It was important that the Shifting Securities website required secured access. For ethical reasons this enabled the project leader to restrict the transcripts and reports to selected guest researchers and maintain ethical fidelity to the informants.

However, there is also considerable resistance and opposition to this developing trend. One objection is based purely on a matter of scale. Dingwall (1992: 169) wrote, 'it is no more possible to reproduce all the data than it is for a film-maker to show every inch of film . . . '. But even selective appendices present problems. There is the issue of confidentiality, and whether or not the identity of informants would be compromised or not. There is also the issue of how the data would be used, and whether it is (or could be) meaningful to people who did not write the notes and do the research themselves. Despite these objections, there is growing pressure among research councils and funding bodies to make at least some forms of ethnographic data more publicly available. But the monograph is still an important source of data in its own right.

Conclusions

Ethnographic methods and participant observation have tended to be overlooked as a research tool in politics and International Relations. However, this is beginning to change and in recent years there has been a renewed interest in political ethnography. Ethnography and participant observation are well-suited to the investigation of political phenomena that other more formal techniques are ill-equipped to examine, such as hard-to-define topics or hard-to-access groups, like guerrilla movements, revolutions, riots, racism, and terrorism.

Participant observation is also well-suited to uncovering what people actually do, particularly when, for whatever reason, what people say they do might be different from what they actually do. If you want to know how a riot broke out, you can ask people afterwards, speak to witnesses, and even speak to participants themselves. But this will provide a different, potentially less reliable source of data, than actually observing first-hand what happens. Participant observation can produce two important sorts of data. The first is to do with what people say. This can either be in the form of answers to questions (like surveys and interviews and focus groups), or revealed statements that come to light in the course of other events. But perhaps the key advantage of participant observation over all other methods is the observational data that it reveals.

However, despite these undoubted strengths of participant observation, it is very difficult to assess the quality of data generated. The problem with participant observation is that it is very difficult to tell whether the data has been collected rigorously or not. We generally only know how long the researcher was in the field, the location of the field site (though sometimes this is anonymized), and the group of people among whom the researcher worked. More methodological appendices would help to establish the reliability and validity of the research, and how far it could be generalized to other contexts. Some of these problems can be overcome. Carrying out comparative ethnography, or replicating existing ethnographies in other contexts, can help to establish the extent to which findings can be generalized. Combining ethnography with other methods of data collection can be used to derive testable hypotheses from the ethnographic data, which can then be examined more generally, using, for example, survey data (see Michelutti and Heath 2013).

Questions

1. What are the particular strengths of ethnographic study? When might it be an appropriate method of analysis? When might it not be so helpful?
2. To what extent does ethnography reflect the ethnographer's own worldview rather than that of the people it purports to represent?
3. Is ethnography only concerned with description? How can it be used to develop explanations?
4. Is ethnography always biased? If so, does it matter? If not, how can we avoid or minimize bias?
5. Are the limits of ethnographic generalization overstated?
6. Should ethnographers deposit their field notes in data archives?
7. What are the ethical obstacles to doing participant observation? Are these an unnecessary inconvenience?
8. How might ethnography and survey data be meaningfully combined?

Guide to Further Reading

Bernard, H. (2006), *Research Methods in Anthropology: Qualitative and Quantitative Methods* (Lanham, MD: Altamira Press).

A very detailed and comprehensive account of the different aspects of ethnography, and the ways in which ethnographic research can be combined with other research strategies and methods. Although written primarily for anthropologists, it is full of engaging examples and is an extremely useful resource for all students interested in political ethnography.

Fenno, R. (1978), *Home Style: House Members in their Districts* (New York: Harper Collins).

This is one of the classic ethnographies in political research. Fenno conducted research with eighteen members of Congress as they travelled around their constituencies, spending time with them, and observing what they did and who they spoke to on a day-to-day basis. The book offers fascinating insights into how Congress people interact with their constituents, and how this translates into their legislative effectiveness.

Hammersley, M. and P. Atkinson (2007), *Ethnography: Principles in Practice* (London: Routledge).

A very clear and comprehensive book on the different aspects of ethnographic research. It contains everything you need to know, with informative and detailed chapters on specific issues such as access, field relations, writing ethnography, and ethics.

Michelutti, L. (2008), *The Vernacularisation of Democracy: Politics, Caste and Religion in India* (London: Routledge).

This is a fascinating ethnography of everyday practices and beliefs about democracy among a historically marginalized caste group in Northern India. The book clearly illustrates what ethnography can contribute to the study of politics. Michelutti examines how culture and social practices inform politics and vice versa. In doing so, she examines how ideas about democracy get ingrained in domains of life, such as marriage, kinship, and religion, which in turn serves to 'vernacularize' democratic politics.

Schatz, E. (ed.) (2009), *Political Ethnography: What Immersion Contributes to the Study of Power* (Chicago: University of Chicago Press).

This book contains a collection of essays on how ethnographic approaches have been used to study a variety of issues in political research, with chapters on topics such as ethnic nationalism, civil war, public opinion, and authoritarianism.

References

Adams, Vincanne (2013), *Markets of Sorrow, Labors of Faith. New Orleans in the Wake of Katrina* (Durham, NC: Duke University Press).

Anderson, Ruben (2014), *Illegality, Inc. Clandestine Migration and the Business of Bordering Europe* (California: University of California Press).

Ashforth, A. (2005), *Witchcraft, Violence, and Democracy in South Africa* (Chicago: University of Chicago Press).

Autesserre, Séverine (2014), *Peaceland. Conflict Resolution and the Everyday Politics of International Intervention* (Cambridge: Cambridge University Press).

Auyero, J. (2006), 'Introductory Note to Politics under the Microscope: Special Issue on Political Ethnography', *Qualitative Sociology* 29(3): 257–9.

Bale, T. and C. Dann (2002), 'Is the Grass Really Greener: The Rationale and Reality of Support Party Status: A Case Study', *Party Politics* 8(3): 349–65.

Barnett, M. (2006), *Eyewitness to a Genocide: The United Nations and Rwanda* (Ithaca, NY: Cornell University Press).

Bernard, H. (2006), *Research Methods in Anthropology: Qualitative and Quantitative Methods* (Lanham, MD: Altamira Press).

Blee, K. and A. Currier (2006), 'How Local Social Movement Groups Handle a Presidential Election', *Qualitative Sociology* 29: 261–80.

Boswell, J., Corbett, J., Dommett, K., Jennings, W., Flinders, M., Rhodes, R., and Wood, M. (2019), 'State of the Field: What Can Political Ethnography Tell Us about Anti- politics and Democratic Disaffection?' *European Journal of Political Research* 58: 56–71

Bratucu, R., Gheorghe, I., Radu, A., and Purcarea, V. (2014), 'The Relevance of Netnography to the Harness of Romanian Health Care Electronic Word-of-mouth', *Journal of Medicine and Life* 7: 363–7.

Brewer, J. (2000), *Ethnography* (Buckingham: Open University Press).

Bryman, A. (2015), *Social Research Methods*, 5th edition (Oxford: Oxford University Press).

Coleman, S. (ed.) (2011), *Multi-Sited Ethnography: Problems and Possibilities in the Translocation of Research Methods* (New York: Routledge).

Comaroff, J. L. (2010), 'The End of Anthropology, Again: Toward a New In/Discipline', *American Anthropologist* 112(4): 524–38.

Comaroff, J. and J. Comaroff (2003), 'Ethnography on an Awkward Scale: Postcolonial Anthropology and the Violence of Abstraction', *Ethnography* 4: 147–79.

Costello, L., McDermott, M.-L., and Wallace, R. (2017), 'Netnography: Range of Practices, Misperceptions, and Missed Opportunities', *International Journal of Qualitative Methods* 16: 1–12.

Dahl, R. (1961), *Who Governs? Democracy and Power in an American City* (New Haven, CT: Yale University Press).

Delamont, S. (2004), 'Ethnography and Participant Observation', in C. Seale, G. Gobo, J. Gubrium, and D. Silverman (eds), *Qualitative Research Practice* (London: Sage Publications), 217–29.

Dingwall, R. (1992), 'Don't Mind Him—He's from Barcelona: Qualitative Methods in Health Studies', in J. Daly, I. MacDonald, and E. Willis (eds), *Researching Health Care: Designs, Dilemmas and Disciplines* (London: Routledge), 161–75.

Eliasoph, N. (1997), *Avoiding Politics: How Americans Produce Apathy in Everyday Life* (Cambridge: Cambridge University Press).

Emerson, R., R. Fretz, and L. Shaw (1995), *Writing Ethnographic Field notes* (Chicago: University of Chicago Press).

Fenno, R. (1978), *Home Style: House Members in their Districts* (New York: Harper Collins).

Fenno, R. (1990), *Watching Politicians: Essays on Participant Observation* (Berkeley, CA: Institute of Governmental Studies, University of California).

George, A. and A. Bennett (2005), *Case Studies and Theory Development in the Social Sciences* (Cambridge, MA: MIT Press).

Gillespie, M. and B. O'Loughlin (2009), 'News Media, Threats and Insecurities: An Ethnographic Approach', *Cambridge Review of International Affairs* 22(4): 667–86.

Gillespie, M., J. Gow, and A. Hoskins (2007), *Shifting Securities: News Cultures Before and Beyond the Iraq Crisis 2003: Full Research Report*, ESRC End of Award Report, RES-223-25-0063 (Swindon: ESRC).

Gurrieri, L. and H. Cherrier (2013), 'Queering Beauty: Fatshionistas in the Fatosphere', *Qualitative Market Research* 16: 276–95.

Hammersley, M. and P. Atkinson (2007), *Ethnography: Principles in Practice* (London: Routledge).

Hannerz, U. (1969), *Soulside: Inquiries in Ghetto Culture and Community* (New York: Columbia University Press).

Hartmann, D. and D. Wood (1990), 'Observational Methods', in A. Bellack, M. Hersen, and A. Kazdin (eds), *International Handbook of Behaviour Modification Therapy* (New York: Plenum), 107–38.

Hopf, T. (2002), *Social Construction of International Politics: Identities and Foreign Policies, Moscow 1955 and 1999* (Ithaca, NY: Cornell University Press).

Hopf, T. (2006), 'Ethnography and Rational Choice in David Laitin: From Equality to Subordination to Absence', *Qualitative Methods* 4(1): 17–20.
Hudson, J. M. and A. Bruckman (2004), 'Go Away: Participan Objections to Being Studied and the Ethics of Chatroom Research', *The Information Society* 20(2): 127–39.
Humphreys, L. (1970), *Tearoom Trade* (London: Duckworth).
Ingold, T. (2008), 'Anthropology is Not Ethnography', *Proceedings of the British Academy* 154: 69–92.
Karlstrom, M. (1996), 'Imagining Democracy: Political Culture and Democratization in Buganda', *Africa* 66: 485–505.
Kozinets, R. (1998), 'On Netnography: Initial Reflections on Consumer Research Investigations of Cyberculture', *Advances in Consumer Research* 25: 366–71.
Kozinets, R. (2010), *Netnography: Doing Ethnographic Research Online* (London: Sage).
Kozinets, R. (2015), *Netnography: Redefined* (London: Sage).
Kozinets, R. V. (2002), 'The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities', *Journal of Marketing Research* 39(1): 61–72.
Kurikko, H. and P. Tuominen (2012), 'Collective Value Creation and Empowerment in an Online Brand Community: A Netnographic Study on LEGO builders', *Technology Innovation Management Review* 2: 12–17.
Lai, D. and R. Roccu (2019), 'Case Study Research and Critical IR: The Case for the Extended Case Methodology', *International Relations* 33(1): 67–87.
Laitin, D. D. (2003), 'The Perestroikan Challenge to Social Science', *Politics and Society* 31(1): 163–84.
Laitin, D. D. (2006), 'Ethnography and/or Rational Choice: A Response from David Laitin', *Qualitative Methods* 4(1): 26–33.
Langer, R. and S. C. Beckman (2005), 'Sensitive Research Topics: Netnography Revisited', *Qualitative Market Research: An International Journal* 8(2): 189–203.
Latour, B. (2010), *The Making of Law: An Ethnography of the Conseil d'Etat* (Cambridge: Polity).
Leach, E. (1961), *Rethinking Anthropology* (London: Athlone Press).
LeBesco, K. (2004), 'Managing Visibility, Intimacy, and Focus in Online Critical Ethnography', in M. D. Johns, S.-L. S. Chen, and G. J. Hall (eds), *Online Social Research: Methods Issues, and Ethics* (New York: Peter Lang).
MacKay, J. and J. Levin (2015), 'Hanging Out in International Politics: Two Kinds of Explanatory Political Ethnography for IR', *International Studies Review* 17(2): 163–88.
Malinowski, B. (1922), 'Introduction: The Subject, Method and Scope of this Inquiry', *Argonauts of the Western Pacific: An Account of Native Enterprise and Adventure in the Archipelagoes of Melanesian New Guinea* (London: Routledge), 2–25.
Marcus, G. (1995), 'Ethnography in/of the World System: The Emergence of Multi-Sited Ethnography', *Annual Review of Anthropology* 24: 95–117.
Michelutti, L. (2008), *The Vernacularisation of Democracy: Politics, Caste and Religion in India* (London: Routledge).
Michelutti, L. and O. Heath (2013), 'The Politics of Entitlement: Affirmative Action and Strategic Voting in Uttar Pradesh, India', *Focaal* 65: 56–67.
Michelutti, L., A. Hoque, N. Martin, D. Picherit, P. Rollier, A. Ruud, and C. Still (2019), *Mafia Raj: The Rule of Bosses in South Asia* (Stanford: Stanford University Press).
Pachirat, T. (2009), 'The Political in Political Ethnography: Dispatches from the Kill Floor', in E. Schatz (ed.), *Political Ethnography: What Immersion Contributes to the Study of Power* (Chicago: University of Chicago Press), 143–62.
Paley, J. (2001), *Marketing Democracy: Power and Social Movements in Post-Dictatorship Chile* (Berkeley, CA: University of California Press).
Parnell, P. and K. Kane (eds) (2003), *Crime's Power: Anthropologists and the Ethnography of Crime* (New York: Palgrave Macmillan).
Parsons, C. (2010), 'Constructivism and Interpretive Theory', in D. Marsh and G. Stoker (eds), *Theory and Methods in Political Science* (Basingstoke: Palgrave Macmillan).
Pieke, F. (1995), 'Witnessing the 1989 Chinese People's Movement', in C. Nordstrom and A. Robben (eds), *Fieldwork Under Fire: Contemporary Studies of Violence and Survival* (Berkeley, CA: University of California Press).
Pouliot, V. (2007), 'Pacification without Collective Identification: Russia and the Transatlantic Security Community in the Post-Cold War Era', *Journal of Peace Research* 44: 605–22.
Rhodes, R. 2011. *Everyday Life in British Government*. Oxford: Oxford University Press.
Riemer, J. (1977), 'Varieties of Opportunistic Research', *Urban Life* 5: 467–77.
Sanjek, R. (1990), 'Ethnography', in A. Barnard and J. Spencer (eds), *Encyclopedia of Social and Cultural Anthropology* (London: Routledge), 243–9.
Schaffer, F. (1998), *Democracy in Translation: Understanding Politics in an Unfamiliar Culture* (Ithaca, NY: Cornell University Press).

Schatz, E. (ed.) (2009), *Political Ethnography: What Immersion Contributes to the Study of Power* (Chicago: University of Chicago Press).
Searing, D. (1994), *Westminster's World: Understanding Political Roles* (Cambridge, MA: Harvard University Press).
Stacey, M., E. Batestone, C. Bell, and A. Murcott (1975), *Power, Persistence and Change: A Second Study of Banbury* (London: Routledge and Kegan Paul).
Tilly, C. (2006), 'Afterword: Political Ethnography as Art and Science', *Qualitative Sociology* 29: 409–12.
Wacquant, L. (2003), 'Ethnografeast: A Progress Report on the Practice and Promise of Ethnography', *Ethnography* 4: 5–14.
Weaver, C. (2008), *Hypocrisy Trap: The World Bank and the Poverty of Reform* (Princeton, NJ: Princeton University Press).
Wedeen, L. (2010), 'Reflections on Ethnographic Work in Political Science', *Annual Review of Political Science* 13: 255–72.
Whiting, J. (1982), 'Standards for Psychocultural Research', in E. Hoebel, R. Currier, and S. Kaiser (eds), *Crisis in Anthropology: View from Spring Hill* (New York: Garland), 155–64.
Whyte, W. ([1943] 1993), *Street Corner Society: The Social Structure of an Italian Slum* (Chicago: Chicago University Press).

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

14 Textual Analysis*

Chapter Summary

Texts have always been a major source of information and evidence for political researchers. This chapter discusses two forms of textual analysis that have become increasingly prominent in Politics and International Relations research: discourse analysis and content analysis. It also offers a brief discussion of using documents, archival sources, and historical writing as data. Among the questions the chapter addresses, are the following:

- What is a discourse?
- What can an analysis of discourse reveal about the social world?
- What is content analysis?
- How do discourse analysis and content analysis differ?
- What are the differences between qualitative and quantitative content analysis?
- What procedures are involved in both quantitative and qualitative content analysis?

Introduction

Government reports, political pamphlets, newspapers, and other texts have long provided important sources of information for students of politics. However, in recent decades, growing awareness of the importance of language and meaning for political analysis, and of the power of the mass media, has produced a dramatic upsurge of interest in textual analysis, not only in Politics and International Relations, but throughout the social sciences. In political research, two forms of textual analysis have become particularly prominent: discourse analysis and content analysis.

Discourse analysis is a qualitative type of analysis that explores the ways in which discourses give legitimacy and meaning to social practices and institutions. *Discourses* consist of ensembles of ideas, concepts, and categories through which meaning is produced and reproduced in a particular historical situation. The elements of a discourse can be brought to light through analysing the language, semiotics (latent meaning in text), and conventions found in a variety of written, oral, and visual 'texts'. But while textual analysis can reveal the elements of a discourse, the *meaning* that they produce or reproduce can only be

understood in relation to some broader context. Consequently, discourse analysis is concerned with analysing, not just the text itself, but the relationship of a text to its context (its source, message, channel, intended audience, connection to other texts and events), as well as the broader relations of power and authority which shape that context.

In contrast to discourse analysis, **content analysis** is concerned with the study of the text itself, rather than with the broad context within which it was produced. This analysis can be either quantitative or qualitative. The aim of *quantitative content analysis* is to draw inferences about the meaning and intention of a text through an analysis of the usage and frequency of words, phrases, and images, and the patterns they form within a text. *Qualitative content analysis* is a more interpretive form of analysis concerned with uncovering meanings, motives, and purposes in textual content.

Qualitative content analysis and discourse analysis have much in common. But a discourse analyst and a content analyst will tend to approach the exploration of an issue in somewhat different ways.

Consider, for example, issues of race and ethnicity, which the upsurge in identity politics (the politicization of ethnic, regional, gender, and sexual identities) has made a key focus of research in recent years. A discourse analyst exploring these issues might examine how ethnic and 'racial' inequality is 'expressed, enacted and reproduced by discourse as one of the practices of a racist society' (van Dijk 2004: 354). In his studies of racism, Teun van Dijk investigates a recurring set of 'underlying mental models and social representations' in 'influential public discourses' (those of elites and elite institutions), and attempts to understand the conditions, consequences, and functions of these discursive elements by exploring the contexts in which they are produced (van Dijk 2002, 2004). A content analyst would be more typically concerned with identifying ethnocentric and racist representations in the content of mass media, literature, and film. For example, in order to find out how the media represented Whites, Blacks, and Latinos as crime victims, Travis Dixon and Daniel Linz (2000) conducted a content analysis of a random sample of television news programmes aired in Los Angeles and Orange Counties. Their examination of the frequency of certain words, phrases, and images revealed that *Whites* are more likely than African-Americans and Latinos to be portrayed as *victims* of crime on television news, and that *Blacks and Latinos* are more likely than Whites to be portrayed as *lawbreakers* than as crime victims. With these data as a basis, the authors were then able to explore the theoretical implications of these findings.

What this example shows is that, despite similarities, discourse analysis and content analysis offer researchers a means of asking and answering different types of questions. In the next section, we discuss these two forms of textual analysis in greater detail, including the types of research questions they address, what they can reveal about the social world, and the procedures that each entails.

Discourse analysis

Discourse analysis is an interpretive and constructivist form of analysis that draws on diverse theoretical and methodological approaches from linguistics, anthropology, and sociology.

As an approach to understanding political phenomena, discourse analysis is interpretive. In common with interpretivist approaches (see Chapter 2), it assumes that people act on

the basis of beliefs, values, or ideologies that give meaning to their actions; and that to understand political behaviour, we must know about the meanings that people attach to what they're doing. Consequently, its aim is to reveal the meanings that the political world has for agents who participate in it and that give people reasons for acting.

But discourse analysis is also constructivist: it assumes not only that people act towards objects, including people, on the basis of the meanings which those objects have for them, but that these meanings are socially and discursively constructed. The aim of discourse analysis, therefore, is not only to reveal meanings through an examination of the language and discourse we employ in our interactions, but to uncover how discursive practices construct meanings through the production, dissemination, and consumption of various forms of texts, including formal written records, TV programmes, advertisements, and novels.

Speech act theory, post-structuralism, and critical discourse analysis

To further elaborate on this constructivist dimension of discourse analysis, it is useful to briefly consider some key ideas on which the constructivist understanding of discourse is based.

Speech act theory

The notion that meanings are discursively constructed is based on a fundamental assumption of discourse analysis: that language is a medium orientated towards action and function; that when we speak (or write), we do so, not only in order to say something, but also to *do* something. This idea—that language is used as much to do things as to make statements—lies at the heart of **speech act theory**, an approach to the explanation of language pioneered by the philosophers John L. Austin and John Searle in the 1960s. 'The issuing of an utterance', wrote John Austin, 'is the performing of an action' (1962: 6); or, as the Austrian philosopher Ludwig Josef Johann Wittgenstein put it: 'Words are deeds' (1980: 46). All discourse-analytic approaches are based on this fundamental idea. Thus, discourse analysis is concerned with language, not as an abstract system, but as something that people use to do things. People use language to promise, threaten, insult, plead, and demand. More broadly, certain things become real through language: 'attitudes', points of view, particular 'subject positions' (ways in which objects or people take up or are placed in particular positions), entities, states of affairs (what is going on), events, processes, activities, and relationships.

So people use *language* to create different kinds of social and cultural meaning, and to construct different accounts, or versions, of the social world. *Discourse* is 'a system of texts that brings objects into being' (Hardy 2001: 26). It constructs social reality by helping to construct what is knowable, sayable, and doable *in a particular historical context*. To highlight the distinction further, we can think of language as a set of rules of grammar and logic that can produce infinite possible sentences (language) or propositions (logic). However, a discourse is a system, or ensemble, of ideas, concepts, and categories, that is linked to, and functions within, a specific context. Consequently, we can say that discourse analysis is an exploration of 'language in context'. It is not interested simply in how something is represented by language, but in how that thing is actually constructed by the ways in which it is intelligible and legitimate to talk about it in a particular time and place.

Post-structuralism

The most important representative of post-structuralist thinking, Michel Foucault (1929–84), focused attention on the various ways discourses 'systematically form the objects of which they speak' (Foucault 1972: 49). This work suggested a variety of channels of inquiry to explore the construction of social reality through discourse, including:

1. how ways of talking about a topic are embedded in sets of power relations;
2. how these power relations are supported by institutions (asylums, governments, prisons, and schools) in particular historical contexts; and
3. how these institutional and historical configurations of discourse constructed new kinds of human subjects.

A particularly resonant avenue of inquiry that Foucault opened up was the exploration of how reiterated key words and statements that recur across texts of all kinds enable and delimit fields of knowledge and inquiry, and govern what can be said, thought, and done within those fields. In his *Archaeology of Knowledge*, Foucault analyses the history or genesis of the 'discursive practices' that lay claim to revealing knowledge, showing how the development of knowledge is intertwined with the mechanisms of (political) power. 'Discourse' can be understood in this context as a system of statements expressing rules, roles, and boundaries that form a body of knowledge. For instance, we can speak of 'discourses of politics', 'discourses of democracy', or 'authoritarian politics'. In *Encountering Development* (1994), Arturo Escobar traces the emergence of a 'development discourse' ('developmentalism') after the Second World War, and shows how this discourse constructed a particular representation of the Third World that, over time, 'created a space in which only certain things could be said and even imagined' (1994: 39).

Post-structuralist theory questions whether there are essential human subjects, individual agents, and social realities independent of their dynamic historical construction in social and cultural discourses. For instance, Foucault asks whether the natural and social worlds are indeed knowable, accessible, and analysable without recourse to the constitutive forces of discourse. He did not think that there were definite underlying structures that could explain the human condition; and he thought that it was impossible to step outside of discourse and survey the situation objectively.

Building on this perspective, post-structuralist discourse theory assumes that the way people talk about the world reflects, not some objective truth about the world, but the success of particular ways of thinking and seeing. These ways of thinking and seeing become invisible because they are assumed to be truthful and right. In this way people's thought processes themselves can come to represent and reinforce particular regimes of power and coercion.

Critical discourse analysis

Post-structural thinking made important contributions to the field of discourse studies. However, it came under attack for its tendency to treat reality as little more than something constituted in and through discourse, and to ignore pre-existing social structures and power relationships. This line of critique was developed in what came to be known as 'critical discourse analysis'.

Critical discourse analysis (CDA) is 'critical' because it seeks to expose connections between language, power, and ideology (Fairclough 2001: 4). It is principally concerned with the role of discourse in enacting, reproducing, and resisting social power abuse, dominance, and inequality (van Dijk 2001: 300, 352).

Hypotheses linking discourses to power can be investigated through an analysis of how powerful groups control public discourse, and of how discourse controls the minds and actions of less powerful groups, and the social consequences of such control (van Dijk 2001: 355).

Discursive power—control over a discourse—is a crucial constituent of social power, and a major means of reproducing dominance and hegemony. Members (and particularly leaders) of more powerful social groups and institutions:

> have more or less exclusive access to, and control over, one or more types of public discourse. Thus, professors control scholarly discourse, teachers educational discourse, journalists media discourse, lawyers legal discourse, and politicians policy and other public political discourse. Those who have more control over more—and more influential—discourse (and more discourse properties) are by that definition also more powerful.
>
> (van Dijk 2001: 356)

Controlling people's minds is another crucial constituent of social power. Those groups who control the most influential discourses also have more chances to control the minds and actions of others, because people tend to accept beliefs, knowledge, and opinions 'through discourse from what they see as authoritative, trustworthy, or credible sources, such as scholars, experts, professionals, or reliable media'. Moreover, 'in some situations participants are obliged to be recipients of discourse, e.g. in education and in many job situations' (van Dijk 2001: 357).

CDA developed in association with a critique of post-structural and other discourse-analytic approaches. But it has also been the target of critique. A key aspect of CDA that distinguishes it from other forms of discourse analysis is the degree to which the analyst depends on a prior theoretical perspective to analyse the data. For CDA, you start with a theory of dominance or power, then you explore materials in order to discover how discourse promotes or challenges one group's power over another (Fairclough 2001). As Ted Hopf points out, critical discourse analysis 'is in fact a political theory as much as a method of inquiry'. It assumes that language is a medium in which prevailing configurations of power are articulated and reproduced. Disrupting and challenging these power relations 'is one of the central features of what we call politics. This means that the meaning of any given text for CDA often points to some underlying political problem or question' (Hopf 2004: 30). This dependence on prior theory brings CDA into conflict with those discourse analysts who want to stay closer to the materials in their data (see e.g. Wetherell 1998; Billig and Schegloff 1999).

Analysis

We have said that a *discourse* is 'a system of texts that brings objects into being' (Hardy 2001: 26) in a particular institutional and historical context. The goal of *discourse analysis* is to explore the relationship between discourse and reality in a particular context. Thus, the hypothesis to be investigated will typically be that that there is a co-variation or association between a discourse and a given context.

To investigate this hypothesis, the analyst will choose a discrete body of written work (e.g. a newspaper over a specific period of time, a set of speeches) and conduct an analysis of what reality its language, metaphors, and/or symbols help to construct. Texts are selected to enable the analyst to explore a hypothesized association between text features and context features, between some change in the context or circumstances and a systematic change in discourse features that are the focus of interest in a study (Lemke 1998). The analysis entails retroduction: reasoning backwards from a particular discursive production to establish structure from its empirical manifestations. 'It asks what the conditions of possibility are of this or that particular discursive production' (Laffey and Weldes 2004: 28). (See Chapter 2.)

Establishing a co-variation or association between discourse and context is only a first step. A second step is to provide details of the process through which the power of a discourse has demonstrable effects.

Let's give some further consideration to each of these two steps.

1. The context

All approaches to discourse assume that, to understand the constructive effects of discourses, researchers must place them in their historical and social contexts. As we stated previously, discourse analysis is the study of language use in context. Discourses construe aspects of the world in inherently selective and reductive ways. The question to ask, therefore, is: 'why this particular selection and reduction, and why here and now?' Thus, discourse analysis 'produces its greatest insights when rich contextual information can be factored into the analysis of each text or episode' (Lemke 1998: 1185). But how do we determine the relevant contextual factors of a text or discourse event? 'Context' is difficult to determine with any precision. It is difficult to know what aspects of context are potentially relevant to a textual analysis.

Let's first consider the different sorts of contexts that can be explored. We can distinguish between *local* and *broad* contexts (Titscher et al. 2000). The *local* context includes the immediate task and situation, the source, message, channel, and intended audience of the communication. The *broad* context consists of cultural norms and assumptions, knowledge, beliefs and values, the resources and strategies characteristic of a community's general cultural resources. Another way of characterizing different contexts is through the use of the terms *micro-discourse* (specific study of language) and *macro-discourse*. 'Language use, discourse, verbal interaction, and communication belong to the micro-level of the social order; power, dominance, and inequality between social groups are typically terms that belong to a macro-level of analysis' (van Dijk 2001: 354)—the broad, societal currents that are affecting the text being studied. Teun van Dijk explains that 'In everyday interaction and experience the macro- and micro-level (and intermediary "mesolevels") form one unified whole. For instance, a racist speech in parliament is a discourse at the microlevel of social interaction in the specific situation of a debate, but at the same time may enact or be a constituent part of legislation or the reproduction of racism at the macrolevel' (van Dijk 2001: 354).

In his studies of discourse and racism, van Dijk focuses on *communicative*, *interactional*, and *societal* contexts. What he is interested in investigating is how 'influential public discourses' (those of elites and elite institutions) reflect 'similar underlying mental models and social representations', and 'similar ways of social interaction, communication, persuasion, and public opinion formation' (van Dijk 2002: 157). To understand the conditions,

consequences, and functions of these discursive elements, he explores the communicative, interactional, and societal contexts in which they are produced. Figure 14.1 represents the way in which these contexts operate.

These contexts together include biased or stereotypical news produced by journalists and other professionals, under the control of editors, in media organizations. In the wider society, the content of these texts is produced, and also reproduced, by members of many different professional and other social groups, and as part of daily routines and procedures: legislative debates conducted by politicians; textbooks, lessons, and scholarly publications produced by teachers and scholars. 'Racist societies and institutions produce racist discourses, and racist discourses reproduce the stereotypes, prejudices and ideologies that are used to defend and legitimize white dominance' (van Dijk 2004: 354).

In a study of how discursive constructions shape the context of policy debates about immigration in the UK, Jonathan Charteris-Black (2006) explores how the British right used metaphors in its political communication to help it gain legitimacy for right-wing positions on immigration policy in the 2005 British election campaign. Using transcripts of speeches given by members of the Conservative Party and party-political manifestos and press articles, Charteris-Black shows that far-right discourses dealing with immigration during the 2005

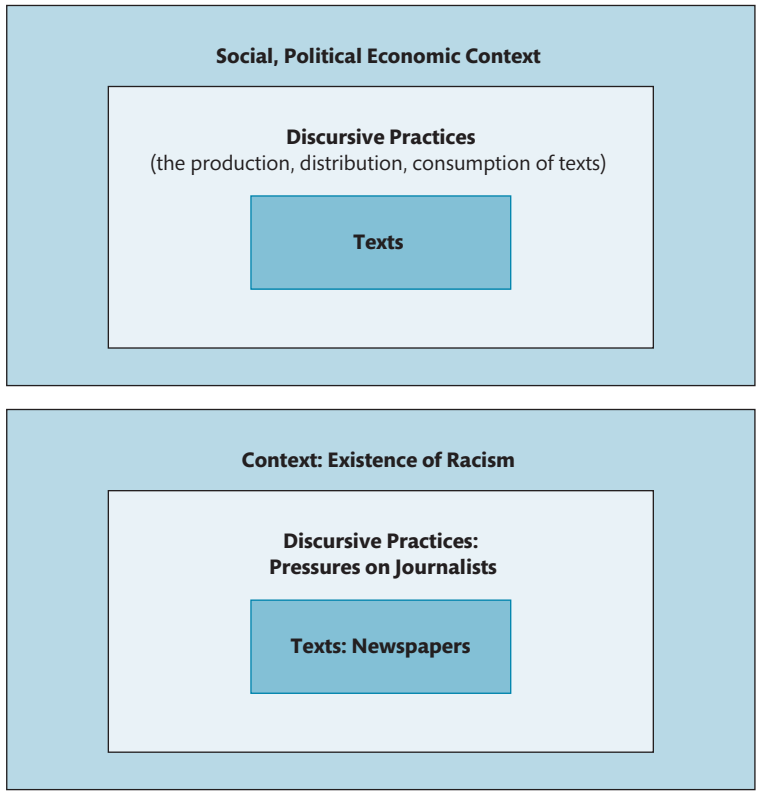


Figure 14.1 Text and context in discourse analysis

Source: Fairclough 1995: 59.

election campaign used two main types of metaphors. The first were the use of 'natural disaster' metaphors in relation to immigration, which alluded, for instance, to incoming 'tides', 'waves', or 'flows' of immigrants. The second were 'container' metaphors, which portrayed Britain as a bounded area protecting what is within from external danger. The use of the container metaphor was meant to imply 'that controlling immigration through maintaining the security of borders . . . will ensure control over the rate of social change in Britain . . . It also heightens emotional fears associated with the penetration of a container' (Charteris-Black 2006: 563). 'Disaster' and 'container' metaphors both arouse fears, and both 'discourage empathy with immigrants by treating them as objects, rather than as the subjects of life stories' (2006: 569).

In another study of how discourse construes aspects of the world, Judith Renner and Alexander Spencer explore how discursive constructions make and unmake 'enemies'. Renner and Spencer note that 'shortly after 9/11 any kind of engagement, let alone reconciliation, with groups associated with al Qaeda (such as the Taliban) was considered impossible and absurd' (2013: 475). However, by 2010, negotiations and reconciliation with the Taliban had become an option (Renner and Spencer 2013: 493). The authors focus on discourse to investigate how this transformation of attitude with the Taliban became possible. They argue that 'the turn to conciliatory strategies became an option as the Taliban were increasingly decoupled from al Qaeda as the primary antagonized enemy of "the West"'. Focusing predominantly on the political speeches by George W. Bush and Barack Obama between 2001 and 2011, they show how:

> the Taliban changed from being understood as 'almost indistinguishable' from the 'terrorist' al Qaeda to being seen as an independent entity worthy of engagement. In addition to this process of decoupling, the signifier 'Taliban' was differentiated step by step into an 'extremist core' on the one hand, which had to be confronted and defeated, and into a group of 'moderate Taliban' on the other hand, with whom reconciliation and reintegration into Afghan society was an option worthy of consideration. (2013: 477)

A study by Rudolf De Cillia, Martin Reisigl, and Ruth Wodak (1999) explores how national identity is shaped by 'discursive strategies of dissimilation (aiming at the construction of national differences) and discursive strategies of assimilation (aiming at the construction of intranational sameness)' (1999: 150). The authors assume that nations can be understood as mental constructs, and that national identities are discursively produced and reproduced. '[T]he idea of a specific national community becomes reality in the realm of convictions and beliefs' through discourses by politicians, intellectuals, and media people that continually construct difference/distinctiveness and uniqueness (1999: 153). The study uses the construction of Austrian national identity as a case study in order to illustrate some prominent linguistic strategies employed to construct nations and national identities, based on an analysis of political speeches, newspaper articles, posters and brochures, interviews, and group discussions.

2. Analysing the process through which the power of a discourse has demonstrable effects

In addition to establishing an association between discourse and context, discourse analysis also seeks to show that discourse has demonstrable effects. In order to do this, discourse analysts have sought to delineate the process through which discourse becomes naturalized so as to become 'common sense'—the mechanisms through which it 'fixes' meanings and

becomes naturalized. These mechanisms can best be understood as a social process with two distinct dimensions: articulation and interpellation (Weldes 1996: 284).

The term 'articulation' refers to 'the process through which meaning is produced out of extant cultural raw materials or linguistic resources' (Weldes 1996: 284). 'Articulation' fixes meanings 'through a process of repeatedly establishing associations between different elements, so that these elements come to be seen as inherently or necessarily connected and the meanings they produce come to seem natural, to be an accurate description of reality' (Weldes 1996: 285).

In an investigation of the process of articulation, the researcher would first identify the main signifying elements of the discourse and how they are articulated to each other. For instance, Stuart Hall has shown how neo-liberal discourses were originally constructed out of a chain of discursive elements linking 'big government', 'unemployment', and 'welfare state', on the one hand, and 'deregulation', 'free markets', and 'privatization', on the other (Hall 1988: 50). Investigation of the process of articulation would, then, focus on how discourses are articulated to institutional forms (think tanks, political parties, and to media outlets) and partake of their power (Hall 1988: 46–7).

In order to be successful, the representations that articulation establishes need to be accompanied by a second dimension of the social process through which discourses 'fix' meanings and become naturalized: the process of 'interpellation'. It is through this process that the acceptance of specific representations is achieved.

What interpellation entails is the acceptance of the 'subject positions' (the placing of people in particular positions) that discourse constructs. Investigating processes of interpellation involves examining how a discourse constructs the identities of subjects and objects, and how these identities are then embodied in the spoken, written, and symbolic texts, the encounters, interactions, and informal talk, of institutional bureaucracies.

An example of an analysis which explores processes through which the power of a discourse has demonstrable effects is Norman Fairclough's study of discourse and social change. Fairclough is concerned to explore the ways in which discourse '(re)constructs' social life in processes of social change (Fairclough 2005a). Discourse, he argues, is a 'crucial and irreducible' element in social change (Fairclough 2005b: 41). To construct an account of how discourse operates as an element of social change, he identifies four 'moments' in the dialectical relationship of discourse and social change (Fairclough 2005b: 42–3):

1. the translating and condensing of complex realities into new discourses constructed through the articulation of elements of existing discourses ('emergence');
2. the contestation among discourses—part of the contestation between strategies and between groups of social agents—which may lead to particular discourses (and strategies) becoming hegemonic ('hegemony');
3. the dissemination of discourses and their recontextualization in new organizations, institutions, or fields, or at new scales ('recontextualization'); and
4. the 'enactment of discourses as new ways of (inter)acting, their inculcation as new ways of being, or identities, their materialization in features of the physical world' ('operationalization').

Fairclough applies this schema to an investigation of the 'discourse moment' in 'the emergence of a new regime of international security and the use of force, as evidenced, in the early 2000s, in Kosovo, Afghanistan, and Iraq': the effort to develop and diffuse a new

hegemonic discourse of international relations and international security. The focus of Fairclough's analysis is the contribution of one key 'player' in this process: the UK Prime Minister, Tony Blair. To explore Blair's contribution to elaborating a new doctrine of 'international community', he analysed '"doctrinal" speeches which elaborate policy' that Blair delivered in April 1999, April 2002, and January 2003 (Fairclough 2005b: 41).

While his analysis is not able to reveal all four 'moments' in the dialectic of discourse and social change, it does suggest what avenues of research this might entail. He finds that the texts he examines reveal, to some extent, the moment of *emergence* of a new discourse through articulating elements of existing ones; and that the second moment, *hegemony*, is revealed inasmuch as Blair is a major international statesman and opinion-former—though, as Fairclough notes, a wider range of material would need to be analysed over a longer period of time 'to get a sense of hegemonic struggles over international relations and international security'. The third moment, *recontextualization*, might be revealed in a further study charting the 'trajectories of the emergent hegemonic discourse in its structural and scalar dissemination'. This might entail 'investigating how the emergent hegemonic discourse of international relations and international security both "colonizes" and is appropriated' within government policy and media texts. He suggests, further, that the operationalization of a change in hegemonic discourse 'in new strategies, institutions, exchanges' might be revealed in case studies of processes of policy formation and implementation by, for example, studying 'the process of decision making and implementation in the procurement of new military hardware' (Fairclough 2005b: 59).

Validity and reliability

Judgements of validity focus on the extent to which a plausible case has been made that patterns in the meaning of texts are constitutive of reality in some way. Does an interpretation adequately account for observations in relation to relevant contextual factors? Does it minimize potential researcher bias, and provide explanatory coherence within a larger theoretical frame?

In addition to plausibility, the validity of a discourse-analytic study can be judged in terms of its credibility—its ability to impart 'coherence' to a text, showing how it fits together in terms of content, functions, and effect; and its 'fruitfulness'—its ability to provide insights that may prove useful (Potter and Wetherell 1994). A discourse-analytic study should demonstrate a careful reading of the text; provide an interpretation that is clearly related to the textual evidence; and present an analysis which aims to be credible, plausible, coherent, and fruitful. Research should be open and transparent, both about the textual evidence under review and about the basis of the claims made about it (Rapley 2008).

Content analysis

Content analysis involves the systematic analysis of textual information. In Politics and International Relations, researchers tend to study election manifestos, news media, and political leaders' speeches. But there is a wide variety of texts that researchers might choose to analyse, including:

(1) *official documents*: government reports and administrative records, programme evaluations, descriptions of programme activities, legal reports, judicial decisions, company accounts; records from schools, hospitals, law courts; judicial decisions; and transcripts of speeches, conversations, discussions, and oral answers to questions;

(2) *cultural documents*: newspaper articles or editorials, magazines, TV programmes, films, videos, art works, photographs, advertising copy; reports from journals, magazines, and newspapers; and

(3) *personal documents*: letters, diaries, and emails.

Content analysis is an unobtrusive method of data collection. Gathering data through unobtrusive means has a number of advantages over obtrusive methods of data collection, such as surveys, unstructured and semi-structured interviews, focus group discussions, ethnography, and participant observation. The chief advantage is that they can reduce bias. For instance, in all types and forms of interviews, people can be expected to come to the interview with biases and prejudices. Moreover, all are generally prone to the 'interview effect': the tendency to give 'socially acceptable' answers or ones that they think the interviewer wants. People may not tell the truth because they want to 'look good' in the eyes of the interviewer, or because they are asked something either that they don't know how to answer or that would be embarrassing for them to answer. Of course, the researcher also comes to the interview with biases and prejudices, and these can distort the interview process as well.

In addition to these problems, obtrusive methods are also prone to the 'Heisenberg Effect', which we discussed in Chapter 3. This is the tendency for people to change their behaviour when they know they are under observation. We described, as an illustration of this, the experience of Charles Frankel, who worked as a United States Assistant Secretary of State in the 1960s. When Frankel sent out letters merely to get information about what US officials were doing with regard to particular programmes, he got back replies indicating that 'the officials to whom he had written had changed what they were doing after receiving his letter' (Morgan 1994: 37). Merely in an effort to inform himself he had 'apparently produced changes in policy', that is, in the phenomena he was studying (Frankel 1969: 83; quoted in Morgan 1994: 37).

Researchers in our field can get around these problems by using unobtrusive methods of data collection. Using content analysis, researchers can get material on decision-making without interviewing the decision-makers. They can systematically analyse an official's statements for evidence concerning their perceptions and attitudes. They can analyse transcripts of a public hearing rather than depend on what government officials remember or choose to tell them about those hearings. In addition to reducing bias, analysing textual information can also enable researchers to gain access to subjects that may be difficult or impossible to research through direct, personal contact; and they can study larger populations and more documents than would be possible through either interviews or direct observation. Research in Action Box 14.1 provides an example of how content analysis has been used by a prominent scholar in our field.

In sum, content analysis might be a way to reduce some types of bias and investigate a wider range of topics among a larger population of people. But whether content analysis is the appropriate analysis for your research project will depend on what evidence you need in order to investigate your hypothesis, and whether the evidence is at least partially embodied in texts.

Research in Action 14.1 National Narratives and Migration

National narratives play a key role in state consolidation and identity construction. All states promote these narratives through official histories. As Laurie Brand notes, 'state elites have an interest in inculcating in the citizenry a particular version of the past', so the content of these histories is carefully considered. Politically sensitive events may likely be reconstructed or completely avoided, and groups who 'do not fit neatly into the state's story may be demonized, rescripted, or simply ignored' (2010: 78). Recently, the difficulties of constructing these narratives have been compounded by large-scale migration and the development of diasporas. These challenge official narratives of national identity and unity and so pose an especially significant problem for the state. Brand observes that state elites respond differently to this challenge: they may completely exclude emigrants from the national narrative, or fully embrace them. So the question she seeks to answer is what governs their choice as to how to present emigrants and emigration in the national narrative?

She starts by identifying four hypotheses about the relationship between a country's emigration history and the way emigrants are treated in its official history. These posit that the treatment of emigrants in official histories will vary depending on their place in a state's colonial history or in its economic development (Hypotheses (H) 1 and 2), and their relationship to state elites (Hypotheses (H) 3 and 4):

H1: a post-colonial regime will tend to *exclude* emigrants if large numbers of emigrants reside in the former colonial state, as this reflects badly on the post-colonial state and its leaders.
H2: a regime may *avoid serious discussion* of emigration when emigration results from its poor performance in economic development and job creation.
H3: state elites will likely chose to *ignore* emigrants if emigrants are not from, or have no special relationship to, them.
H4: state elites will likely *include* immigrants if they represent a source of power to elites, and *exclude* them if they represent a threat to elites.

Brand (2010) then selects, as case studies, two countries which have experienced significant emigration: Jordan and Lebanon. After exploring the place of emigrants in their history, she then undertakes a detailed examination of 'the most obvious, concrete, and reliable source of the state-sanctioned national story': government history and civics textbooks. These carefully crafted texts represent the most effective means by which state-sanctioned historical narratives are communicated to citizens: they are used by all children in public educational institutions and expose children to the national narrative repeatedly over years.

Her historical research shows that all of Lebanon's communal groups have contributed to Lebanon's large diaspora and all have a sense of pride in its achievements. It provides a basis of national identity and unity and thus a source of power for state elites. Jordan's emigration history, however, is politically sensitive and has severely challenged its unity. Large numbers of its Palestinian population see Jordanian policies towards Palestine or the Palestinians as problematic or even treasonous. The sensitivities generated by Palestinian refugee movements caused by war, have been compounded by large-scale Palestinian migration to the Persian Gulf as a result of policies which have excluded Palestinians from job opportunities in Jordan.

After detailed examination of school textbooks used in the two countries, Brand (2010) then revisits her initial hypotheses. H1 is inapplicable: there was no Jordanian emigration to Great Britain; and the Lebanese community in France is not a result of French colonial era policies. H3 is inapplicable because emigrants are included in both countries' official histories. H2 has some relevance. Jordan's labour emigration, a function, in part, of Jordan's underperforming economy, receives scattered treatment in textbooks, but it is not attributed to government policies. Lebanese textbooks offer a more thorough and forthright discussion of emigration, though they attribute it to external forces rather than to the Lebanese or the government.

(Continued)

H4 provides the best explanation of both cases. Lebanon, where emigrants represent a source of power to elites, exposes students early and consistently to the importance of emigration in the country's history; and textbooks extol the Lebanese diaspora for its initiative and contributions to the country, rather than demonizing or excluding it. In Jordan, where emigrants represent a threat to elites, schoolbooks tend to suppress or exclude them. Brand concludes that these cases show that

> [i]f the emigrant/migrant experience reinforces the legitimacy of the existing power structure or can be scripted to perform such a function . . . the regime may seek discursively to instrumentalize the migrant experience to its benefit. If, on the other hand, the country's migration history poses challenges to the regime's claims to rule or its hold on power . . . the narrative may be constructed to marginalize or even ignore the migrant experience.
>
> (Brand 2010: 108–9).

Source: Brand (2010).

Qualitative and quantitative content analysis

Content analysis can be either quantitative or qualitative. We tend to agree with Klaus Krippendorf, that '[u]ltimately, all reading of texts is qualitative, even when certain characteristics of a text are later converted into numbers' (2004: 16). However, quantitative and qualitative content analysis differ in the types of questions they address, as well as in the procedures they use both to analyse text and to record, process, and report data. Which method you employ will depend on your research question and the hypothesis you intend to investigate.

The development of *quantitative content analysis* was inspired 'by the need to develop a more objective and systematic method for analysing the rapidly increasing volume of communications produced by governments, companies, and other organizations' (Burnham et al. 2004: 236). It has generally been concerned with the **manifest content** of communications. This is the content that is easily observable—that resides on the surface of communication. An example would be to count the number of times a particular word, phrase, or image occurs in a communication. Quantitative analyses are generally concerned to tell us 'How often?' or 'How many?'

However, not all research questions can be answered by focusing on the manifest or surface content of texts. The meaning of the text, or the variables of interest to the researcher, may not reside on the surface of the content but 'between the lines'. *Qualitative content analysis* is more concerned with this **latent content**. Qualitative content analysis assumes that it is possible to expose the meanings, motives, and purposes embedded within the text, and to infer valid hidden or underlying meanings of interest to the researcher (Weber 1990: 72–6). It is generally more sensitive to the context in which texts are produced, and better able to tell us about meanings, norms, values, motives, and purposes. However, as we shall discuss later in this chapter, recent advances in computer-aided content analysis have also enabled quantitative analyses of content to uncover latent dimensions of texts.

Irrespective of whether you choose to conduct a quantitative or a qualitative content analysis, you will need to address the same considerations before embarking on a content

analysis. You will need to decide what population of documents would provide evidence relevant to your hypothesis, what content you will examine and why, what kinds of data are required, whether they are available or need to be collected, the kind of analysis that is required, and the resources needed. And, irrespective of whether your analysis is quantitative or qualitative, the procedures that you will need to follow to collect data are generally the same.

Steps to carrying out content analysis

Quantitative and qualitative content analysis generally involves the same four steps. We can think of each of these steps as relating to a specific question that the researcher must ask and answer.

Step one

Question: What set of documents is germane to your research question, and what sample from this set will you analyse?

Answer: Select both the population of texts you will use and how much of this material is to be analysed.

The first step is to *select the material to be analysed*. This requires that you first *identify the population of texts* (documents or other communications) that will provide evidence appropriate to an investigation of your hypothesis. Once the population of relevant texts has been identified, you will need to be sure that all of the documents that you need from this population are available and accessible. If some cannot be located—if, for instance, there are missing years in a series of annual reports that you want to investigate—you may risk introducing bias into your analysis. Once you have identified the population of relevant documents and ensured that they are available, you must decide whether to analyse the full set or a partial set of the material. If the document population is too large to be analysed in its entirety, you will need to *select a representative sample* of the material to investigate and analyse. Probability sampling may be the right choice if your intention is to generalize from the sample to the population. Non-probability sampling can be used if generalization is not necessary or if probability sampling procedures are not practical.

Depending on the aims of the analysis, you may choose textual data that belongs to a single semantic domain, for example political speeches, or that cuts across a number of semantic domains. Your data might consist of one particular text type (party manifestos, annual reports) or include texts belonging to different text types. The texts you select might contain only texts produced by a specific speaker or author, either of the same, or different, types; or they may be produced by different persons—as in the case of open responses to questions in surveys. The data can be obtained from an existing archive, or they may need to be collected. There exist a large number of electronic text archives containing text data from a great variety of sources, and online text databases with a wide variety of text material which can be directly accessed and downloaded. In addition, full texts of a variety of publications are available online.

Step two

Question: What will you examine these texts *for*?
Answer: Define categories.

Once you have selected both the population of texts you will use and how much of this material you will analyse, the second step is to *define the categories* or topics of interest that you will search for in the material you have selected to analyse. For a quantitative content analysis, this requires you to be clear about the variables you want information about: the subjects, things, or events that vary and that help to answer your research question. Then you must identify categories for each variable. For example, you might identify three categories related to the variable 'attitude towards the war': negative, neutral, and positive. Or you might define five categories for the variable 'attitude towards the incumbent': greatly dislike, moderately dislike, indifferent to, moderately like, and greatly like.

Step three

Question: What segments of the text will contain what you are searching for*?*
Answer: Choose the recording unit.

The third step is to *choose the recording unit* (unit of content)—the portion or segment of text to which you will apply a category label. There are five recording units normally used in content-analytic studies:

1. A *single word or symbol*. This is generally the smallest unit that is used in content analysis research, and it is an appropriate unit if you are studying the use of language. When words are the recording unit, researchers are usually interested in investigating the choice of certain words or counting the frequency with which they occur. *Word sense*—words that convey values (positive, negative, indifferent) or ideological positions—is a variation on words as recording units; and they can be counted just as if they were words.
2. A *sentence or paragraph*.
3. A *theme*. The boundary of a theme delineates a single idea, or 'a single assertion about some subject' (Holsti 1969: 116). This might be the recording unit in research on propaganda, values, or attitudes and beliefs.
4. A *character*. This might be the recording unit in studies of fiction, drama, movies, radio, and other forms of entertainment materials.
5. An *item or whole text*. This might be an entire article, film, book, or radio programme; a newspaper item, a magazine article, a web page or book, an episode of a TV programme, or a transcript of a radio interview.

Quantitative content analysts are generally more likely to divide texts into smaller segments: to count individual words, or examine phrases, word-strings, sentences, or paragraphs. Qualitative analysts tend to study documents in their entirety. In all cases, selection of recording units must be based upon the nature of the variables and the textual material to be coded. However, trade-offs among different options should also be considered. A recording unit that is objectively identifiable—one that has obvious physical (e.g. a whole

text) or semantic boundaries (paragraphs, sentences, or words)—makes the coder's task relatively easy. However, while a recording unit that is objectively identifiable has its advantages, it may not properly encompass the categories being investigated. Sometimes a paragraph embraces too many ideas for there to be consistent assignment of the text segment to a single category. In fact, the larger the size of the recording unit, the more difficult and subjective is the work of coding, because as the size of the unit expands, so does the likelihood that the unit will encompass multiple variables.

Step four

Question: How will you identify, and signal the presence in your recording units of, the categories you are looking for?

Answer: Create (a) a coding protocol; and (b) a code for each variable, or a tag for each theme or topic, you are looking for; and (c) mark the text with the codes or tags.

Once these steps are completed, you are ready for the fourth step: **coding**. This involves (a) *creating a protocol* for identifying the target variables and categories; (b) *creating codes* that will signal their presence in the text; and (c) *coding* the texts using the protocol and codes.

Creating a protocol (a) involves developing a set of rules to ensure the reliability of the coding. The protocol will reflect a set of decisions that ensures that the researcher will code things consistently throughout the text, in the same way every time. For instance, the researcher has to decide whether to code only a predefined set of categories, or whether relevant categories not included in the set can be added as they are found in the text. Using a predetermined number and set of concepts allows the researcher to examine a text for very specific things. On the other hand, providing for flexibility allows new, important material to be incorporated into the coding process that could be significant for the findings of the analysis. Much quantitative analysis uses predetermined categories, though advances in computer technology (to be discussed later) have made it possible for quantitative analysis to uncover latent categories, rather than using predefined categories. Most qualitative content analysts prefer not to pre-set categories but to allow categories to emerge out of the data (Bryman 2004: 183).

Another decision is whether to code only for a given word or also for words that imply that word. Can words be recorded as the same when they appear in different forms? For example, 'agree' might also appear as 'agreement'. The researcher needs to determine whether two words are similar enough that they can be coded as being the same thing, that is, 'agreeing words'. Researchers concerned to draw conclusions about the importance of a topic in the print or broadcast media might measure newspaper space (column inches) or radio or television airtime.

These two decisions might best be understood as completing a sequence of steps in which you (1) ensure your analysis will be focused on the variable of interest to you (Step one); (2) determine the possible values each variable can take (Step two); and (3) provide operational definitions of the variable's values that specify what phenomena you must observe to identify its existence (Step three).

Once you have created a protocol, you can then *create a* ***code*** (b), or short tag, for each variable's categories. Codes are simply abbreviations, or tags, for segments of text. Typically, a code will be an abbreviated version of a category. For example, a researcher coding three categories related to the variable 'attitude towards the war'—negative, neutral, and positive—might label these categories 'attwar*n*' (for negative), 'attwar*0*' (for neutral), and 'attwar*p*' (for positive).

The final step is *coding* (c): marking recording units with the appropriate tags that you've designed to identify the categories being sought in the text. (Before you start, make sure each document from the set of working documents is recorded in a log and given a unique number. As the coding proceeds, record additional information, such as the name of the coder, the date it was coded, and unusual problems.)

Coding

Coding involves the identification of passages of text (or other meaningful phenomena, such as parts of images) and applying labels to them that indicate they are examples of some thematic idea. The coding process enables researchers quickly to retrieve and collect together all the text and other data that they have associated with some thematic idea so that they can be examined together, and different cases can be compared. We can distinguish between two broad approaches to coding: those using a priori codes, and those relying on grounded codes. A priori codes are based on a research hypothesis or a range of sources relating to it, such as previous research or theory, or topics from your interview schedule. Relying on a priori codes is what is often referred to as 'closed coding'. Grounded codes emerge from the data as the researcher reads it. The researcher puts aside presuppositions and previous knowledge of the subject area and concentrates instead on finding themes in the data. This is often referred to as 'open coding'. The most common way to go about developing grounded codes is through 'constant comparison'. What this means is that every time a researcher selects a passage of text and codes it, it is compared with all the passages that have already been coded that way. The researcher asks: 'what is this about?' and 'how does it differ from the preceding or following statements?' A list of the codes is drawn up, with a short definition attached to each one. Each time you find a passage that appears as though it might be coded with an existing code, you can check the coding frame or list to be sure that the passage fits with the definition. If there isn't an appropriate code, or the text doesn't fit with the definitions, then you can create a new one. Eventually, you may want to sort codes into groups. You may find that several codes group together as types or kinds of something; that they refer to different ways that people react to, categorize, or cause something and so might be seen as dimensions of that thing (Strauss and Corbin 1990).

Of course you can move from one coding approach or strategy to another. For instance, some researchers suggest a three-stage coding process. In the first, 'open coding' phase, the researcher carefully reviews a small sample of the documents, making general notes about the broad themes that characterize each individual document, as well as the entire set of texts. In the second stage, all the documents are reviewed with these themes in mind. Patterns are labelled, and passages are 'tagged' as belonging to one or more categories. In the third stage, these labels and tags are checked and re-checked to ensure that they are applied properly (Neuman and Robson 2007: 337–42).

Manual and computer-assisted coding

Textual material can be coded manually, that is, reading through the text and manually writing down concept occurrences; or it can be coded on the computer through the use of various computer programs.

In manual coding, the coder transcribes notes and interviews and copies transcripts and images; they then make multiple copies of everything, as each item may represent an example of more than one theme or analytic idea. In open coding, perhaps using coloured highlighters, the coder marks the text by circling or underlining words, or running lines down the margins to indicate different meanings. In closed coding, the coder simply marks the boundaries of the recording unit and writes the code in the margin of the document, perhaps using different coloured pens for each variable. Next, the coder will cut up the transcripts and collect all the text fragments that are examples of similar themes or analytic ideas (open coding), or that are coded the same way (closed coding), into piles, envelopes, or folders. Each group of fragments can then be set out and re-read to discover more specific patterns or features of significance.

Coding can also be performed automatically with computer software that codes specific parts of text according to the particular categorization scheme you construct. A large number and variety of increasingly user-friendly computer software programs have been developed for this purpose.

Quantitative content analysis programs allow you to examine very large amounts of data, and a wide range of texts, quickly and efficiently. Among the widely used programs for quantitative content analysis are: General Inquirer, KH Coder, Wordsmith, Textpack, TACT (Text, Analysis Computing Tools, and Textstat—a freeware program for the analysis of texts). There are also software programs that have been developed specifically for qualitative content analysis, including ATLAS.ti, MAXQDA, and Qualtrics. These allow the researcher to identify the recording unit in the text and assign the text to a coding category that has been defined either in advance or in the analysis process. They allow multiple coding of individual passages, and also allow multiple coders to work on a single coding task while maintaining identification of the coder for calculation of reliability. A wide variety of reports can be generated from these packages, including counts of codes with illustrative quotations from the text.

Computer-assisted qualitative data analysis software (CAQDAS) allows researchers to use computer-based directories and files rather than physical files and folders, and to use word processors to annotate texts. CAQDAS packages, like QSR NVivo, help to organize and analyse data in documents, as well as in pictures, audio, and video.

Dictionary-based content analysis programs provide a 'basic handful' of text analysis functions, including word and category frequency counting and analysis, sorting, and simple statistical tests. 'Word frequency' refers to how often each word occurs in a document. Most operating systems (Windows, Mac OSX, Unix/Linux) have utilities to perform basic word counting and sorting. But software packages allow you to exclude 'stopwords', common words like 'in' and 'the' which add little meaning but get in the way of the analysis. They can produce a key words in context (KWIC) concordance—a list of the principal words in the document, in alphabetical order, and their immediate context. They also include lemmatization, which involves combining all words with the same stem, such as intend, intended, intends, intending, intent, intention, intentions, etc. For *category frequencies*, programs will group synonyms into categories and then show how many times each category occurs in the document. For example, the Linguistic Inquiry and Word Count (LIWC) dictionary maps onto its category 59, 'death', the following words: ashes, burial*, buried, bury, casket*, cemet*, coffin*, cremat*, dead, death*, decay*, decease*, deteriorat*, die, died, dies, drown*, dying, fatal, funeral*, grave*, grief, griev*, kill*, mortal*, mourn*, murder*, suicid*,

terminat*. The asterisks tell the program to treat all words matching a stem or stem word as belonging to the same category (Lowe 2007: 2). By collecting into a single category different words or phrases that represent a concept, category counts helps to make manifest the latent content in texts.

Another set of programs contains *annotation aids* (e.g. QDA Miner; QSR NUD*IST; ATLAS.ti). These consist of an electronic version of the set of marginal notes and cross-references that researchers use when marking up transcripts by hand in order to analyse them and discover patterns.

An instructive example of the purposes for which researchers use the systematic analysis of political texts is the work of the Comparative Manifestos Project (CMP). The CMP is a large and influential quantitative content analysis project concerned with measuring the policy positions of political parties. The CMP produced quantitative content analyses of parties' election programmes from fifty-one parliamentary democracies covering all democratic elections since 1945. This has become a widely used data set for party positions.

The CMP (previously known as the Manifesto Research Group (MRG)) was formed in 1979.[1] The CMP undertook to measure the political preferences of parties across time and space through a comparative content analysis of political parties' election manifestos. Party manifestos are the programmes political parties issue for elections in some parliamentary democracies. They are authoritative party policy statements which set out both their strategic direction and the legislative proposals they intend to pursue should they win sufficient support to serve in government. The political researchers involved in the project were interested in addressing two questions: (1) what political issues divided post-war political parties; and (2) were they converging or diverging in ideological or policy terms? Anthony Downs' influential model of two-party competition leads to the expectation of party convergence to the policy position espoused by the median voter (Downs 1957: 112–19). MRG researchers were concerned to find out whether this actually occurred.

To estimate the policy position of a particular party on a particular matter at a particular election, the CMP used a trained human coder to manually code the party's election manifesto.[2] First, a classification scheme was developed consisting of fifty-seven policy-coding categories. The unit of analysis that was used was a 'quasi-sentence'. A quasi-sentence is defined as 'an argument or phrase which is the verbal expression of one idea or meaning' (Klingemann et al. 2006: xxiii). Since long sentences may contain more than one argument, breaking up text into quasi-sentences enables researchers to isolate individual ideas so that they can be analysed. Coding consisted of allocating every quasi-sentence contained in a party's manifesto to one, and only one, of the fifty-seven categories (one of which was 'uncoded'). A left–right position measure was calculated by grouping issue categories into 'right' and 'left' categories and subtracting one from the other.

Once text units were allocated to each category and counted, the CMP then defined the relative salience for the party of the policy area defined by each category as the percentage of all text units allocated to that category. The coding scheme derived from the documents themselves. Researchers found that relative emphases were the way in which British parties expressed themselves. Parties compete with each other by emphasizing different policy priorities, rather than by directly opposing each other on the same issues. For instance, political opposition was expressed by emphasizing peace, as opposed to military strength, freedom as opposed to planning. So the analysis was not concerned with counting *positive*

or *negative* references made to different policy areas, but the *relative emphasis* parties placed on these categories.

The CMP data were generated by party manifestos coded once, and once only, by a single human coder. Though manual coding has advantages, the coding protocol had disadvantages. The manual analysis of party manifestos is an extremely time-consuming and costly process. Moreover, there is a potential for coding bias because human coders are inevitably aware of the authorship of the texts they are coding. These flaws in its manual coding encouraged CMP researchers to develop computer programs for the content analysis of party positions. A major step forward was the development of a fully automated text analysis program called Wordscores for measuring policy positions of texts (Laver, Benoit, and Garry 2003). Wordscores uses reference texts and reference values in order to predict policy positions. The basic idea is that we can estimate policy positions by comparing two sets of texts: 'reference texts', documents for which we know the policy positions (e.g. by relying on expert surveys), and 'virgin texts', documents about which we do not know anything apart from the words they contain. A more recent innovation in quantitative content analysis of party positions, Wordfish, is a program that uses a statistical model of word counts, rather than anchoring documents. Wordfish estimates the policy positions of political actors based on the assumption that words are distributed according to a *poisson distribution* (the probability of a number of events occurring in a specified interval) (Proksch and Slapin 2009).

Computerized coding is easier, more flexible, less costly, and less labour intensive; and by removing the human factor from the coding process, it can significantly enhance the reliability of the content analysis. Thus, efforts to develop programs that resolve shortcomings in these methods are continuing; and because of the salience of the work generated by the CMP, the innovations of CMP researchers in this area have become a primary impetus behind the move to computerize content analysis in political research more generally.

Analysis

Analysis is the process of making sense of and attaching meaning to the data we have gathered, and applying the resulting knowledge to our research question.

Once the coding is done, the researcher examines the data for patterns and insights relevant to the key research issues. The codes might be combined or sorted into families for more meaningful analysis. The data is analysed either to describe the target variable(s), or to identify themes or relationships between variables. The researcher then attempts to draw conclusions and generalizations by linking the data back to the research question. Irrespective of whether the data are quantitative or qualitative, the analysis will involve an attempt to inferentially link the textual data to the specific events, behaviour, or phenomena that are of interest to the researcher.

In a quantitative content analysis, analysis involves examining numerical data in relation to pre-operationalized variables, and drawing inferences based on the frequency, amount, salience, or intensity of a category (e.g. the intensity of a person's opinions or attitudes). The inferences drawn will be based on what resides on the surface of communication, that is, what has been said, and on the observation of patterns that are explicitly present, but which may have been hard to see just by reading all the texts. The researcher may also be able to discover relationships among those patterns, and to identify much larger trends or ideas

from the patterns; and, if the research involves charting changing language or other textual features over time, the researcher can draw inferences about political change. Qualitative data is non-numerical (words, images). Its analysis is conceptual, and involves identifying themes that emerge from the data. Quantitative content analysts report their data largely in numeric form, be it statistical, graphical, tabular, or figural. They may use frequency scores to generate 'word clouds' (or 'tag clouds'), which use the frequency of words in a given piece of text to generate a visual representation of the document. Qualitative content analysts rely on quotations and narrative as their primary modes of presentation. Some also draw concept maps, charts, diagrams, or other figures to visually represent the patterns in their data. The differences in how quantitative and qualitative researchers typically conduct a content analysis are summarized in Box 14.1.

Validity and reliability

As with any method of data collection, researchers using content analysis must be concerned with the validity of their analysis and the reliability of their results.

In quantitative studies, validity refers to the extent to which we can draw unambiguous conclusions from our results, and whether our conclusions are likely to apply to other similar situations or cases. A study is valid if its measures actually measure what you claim they measure, and if your inferences follow from the data. Reliability refers to the 'repeatability' or 'consistency' of your findings. A study is reliable if anyone else following the same procedures

BOX 14.1 Quantitative and Qualitative Content Analysis

	Quantitative	Qualitative
Objects of observation	Manifest content: word usage, sequences	Latent content: meanings, motives, purposes
Recording units (units of observation)	Segments of text	Whole texts
Procedures of observation	Counting, rating, logging	Theming, tagging, memoing
Discovery of patterns	Calculated during analysis	Developed throughout process
Presentation of data	Graphs, tables, statistics, figures, word clouds	Quotations, concept maps, narrative
Data format	Numerical (frequency, amount, salience, intensity)	Non-numerical (words, images)
Data reduction	Variables (operationalized a priori)	Themes (emergent)
Substance of data	Meaning is inherent	Meaning is contingent
Data recording	Standardized instrument	Variable instrument
Data processing	Mathematical	Conceptual
Data reporting	Statistical, graphical	Verbal
Standards of evidence	Probability	Plausibility

Source: Adapted from Wesley (2011: table 6.2).

would get the same results. The reliability of a content analysis study depends on three elements. The first is *coder stability*: does the same coder consistently recode the same data in the same way over a period of time? Second, is *reproducibility*: do the coding schemes lead to the same text being coded in the same category by two or more coders? Different people should code the same text in the same way. The simplest and most common method of reporting intercoder reliability is the *per cent agreement* statistic. This statistic reflects the number of agreements per total number of coding decisions. 'Percent agreement after discussion' refers to reliability figures that were obtained through discussion between coders. Holsti's (1969) coefficient of reliability (CR) provides a formula for calculating per cent agreement:

CR = 2m/n1 + n2,

where:

m = the number of coding decisions upon which the two coders agree
n1 = number of coding decisions made by rater 1
n2 = number of coding decisions made by rater 2.

The third, and final, element on which reliability of a content analysis study depends is *objectivity*. Intercoder reliability reveals objectivity by showing the extent to which different coders, each coding the same content, come to the same coding decisions (Rourke et al. 2001: 13).

There are a number of common sources of unreliability:

1. A document is poorly written or vague;
2. Word meanings, category definitions, or other rules in the coding instructions are ambiguous;
3. There is a lack of objectivity in the process of category definition and in the coding of sections of text;
4. The coder makes mistakes.

In qualitative analysis, the results of a study are valid and reliable to the degree that they are plausible to others: that is, if the researcher explains how they came up with the analysis in a way that the reader can make sense of.

Both quantitative and qualitative content analysts should make both their data and raw materials available for verification (i.e. coding databases, memos, and the original documents). In quantitative content analysis, which uses a standardized coding instrument, this is most efficiently accomplished through the publication of the coding manual, including a comprehensive list of coding rules. Qualitative analysts must provide their readers with a detailed account of the coding 'protocol', including how conclusions were reached (Altheide 1996: 25–33). As Holliday suggests, all research 'needs to be accompanied by accounts of how it was really done . . . [Analysts must] reveal how they negotiated complex procedures to deal with the "messy" reality of the scenarios being studied' (2007: 7).

Box 14.2 reviews some key terms introduced in this chapter.

Unlike asking people questions (e.g. through surveys or in interviews), using texts to collect data has the advantage of being non-intrusive. Researchers do not face the problem of influencing their data source through the questions they ask, and they can study past policy positions as they were recorded at the time. Once recorded, texts do not change.

BOX 14.2 Key Terms

Discourse analysis	A qualitative type of analysis that explores the ways in which discourses give legitimacy and meaning to social practices and institutions.
Critical discourse analysis	An approach that assumes that discursive power—control over a discourse—is a crucial constituent of social power, and a major means of reproducing dominance and hegemony; and seeks to expose connections between language, power, and ideology.
Speech act theory	An approach to the explanation of language premised on the idea that language is a medium orientated towards action and function; that language is used as much to do things as to make statements.
Post-structuralism	An approach concerned with how discourses are embedded in sets of power relations and supported by institutional and historical configurations; and, particularly, how the 'discursive practices' that lay claim to revealing knowledge are intertwined with the mechanisms of (political) power.
Content analysis	Analysis of the content of a text in order to uncover its meanings and intentions.

Conclusions

This chapter has outlined two main forms of textual analysis—discourse analysis and content analysis. Each of these forms of textual analysis provides insights into political phenomena: discourse analysis through examining how discursive practices construct the identities of subjects and objects and exercise power; qualitative content analysis through exploring the meanings, motives, and purposes of political action embedded within texts; and quantitative content analysis by drawing inferences about opinions or attitudes from an analysis of the usage and frequency of words, phrases, and images, and the patterns they form within a text. Despite their very real differences, similar standards apply to both forms of textual analysis. Both must be trustworthy in their treatment of documents. Both must be concerned with the validity and reliability of their procedures and conclusions.

Box 14.3 summarizes the differences and similarities among these forms and approaches to textual analysis.

Which of the two forms of textual analysis is 'better' than the others can only be determined in relation to a specific research project, and will depend on the research question, what sort of analysis will provide a useful response to it, and what data are needed and from whom. As we have previously emphasized, techniques of data collection are always employed in the service of a research question. As with any method of data collection, the use of discourse and content analysis should be appropriate to the research question and hypothesis you are investigating. The data collected by these means do not 'speak for themselves'. They are only interesting and significant to the extent that they provide a means of investigating a research question and hypothesis. You choose these means of data collection because you are confident that the data that they provide will enable you to investigate and draw logical conclusions about your hypothesis.

BOX 14.3 Discourse Analysis and Content Analysis Compared

	Discourse analysis	Content analysis	
		Qualitative	Quantitative
Ontology	Constructionist—assumes that reality is socially constructed	There is no inherent meaning in the text; meanings are constructed in a particular context; and the author, consumer, and researcher all play a role	Realist—assumes that an independent reality exists
Epistemology	Meaning is fluid and constructs reality in ways that can be posited through the use of interpretive methods		Meaning is fixed and reflects reality in ways that can be ascertained through the use of scientific methods
Data source	Textual meaning, usually in relation to other texts, as well as practices of production, dissemination, and consumption		Textual content in comparison to other texts, for example over time
Method	Qualitative (although can involve counting)	Qualitative (although can involve counting)	Quantitative
Categories	Exploration of how participants actively construct categories	Categories emerge from the data, though the research question, existing empirical research, and theory provide ideas for what to look for	Analytical categories taken for granted and data allocated to them
Inductive/deductive	Inductive	Inductive	Deductive
Subjectivity/objectivity	Subjective	Subjective	Objective
Role of context	Texts can be understood only in discursive context	Can only understand texts in discursive context	Does not necessarily link text to context
Reliability	Formal measures are not a factor, but coding is justified according to academic norms; different interpretations are not a problem and may be a source of data	The results are reliable to the degree that they are plausible to others: that is, does the researcher explain how they came up with the analysis in a way that the reader can make sense of?	Formal measures of intercoder reliability are crucial for measurement purposes; differences in interpretation are problematic and risk nullifying any results

(Continued)

	Discourse analysis	Content analysis	
		Qualitative	Quantitative
Validity	Validity is in the form of demonstrating a plausible case that patterns in the meaning of texts are constitutive of reality in some way	The results are valid to the degree that they show how patterns in the meaning of texts are constitutive of reality	Validity is in the form of accuracy and precision—demonstrating that patterns in the texts are accurately measured and reflect reality
Reflexivity	Necessarily high—author is part of the process whereby meaning is constructed	Considers the extent to which the author plays a role in making meaning, and different ways a meaning might be consumed	Not necessarily high—author simply reports on objective findings

Source: Adapted from Hardy, Harley, and Phillips (2004: 20–1).

Questions

1. What do discourse and content analysis offer to political analysis?
2. How do you know when you've identified a discourse? Where are its boundaries? Do discourses overlap? In what context or set of conditions does a discourse exist?
3. How can you evaluate whether, when, and how political texts have effects on political life?
4. With what general type of research questions is content analysis concerned?
5. What is the difference between manifest and latent content? What are the implications of this distinction for content analysis?
6. If politics is about power, and language has power, is political analysis a matter of analysing language?
7. In what ways might your analysis address the issue of selection bias when using historical writing as data?

Guide to Further Reading

Discourse analysis

Clayman, S. and J. Heritage (2002), *The News Interview: Journalists and Public Figures on the Air* (Cambridge: Cambridge University Press).

A comprehensive source for information about news interviews, beginning with the historical development of the news interview in the United States and Britain, the norms and conventions that shape interview practices used by journalists, and the use of well-known interviews to explore the relationship between journalists and public figures.

Fairclough, N. (2002), *Analysing Discourse: Textual Analysis for Social Research* (London: Routledge).

An introduction to discourse analysis, drawing on a variety of texts—from political speeches and television news reports to management consultancy reports and texts concerning globalization—to illustrate key issues in discourse analysis.

Gee, J. P. (2005), *An Introduction to Discourse Analysis: Theory and Method*, 2nd edition (New York: Routledge).
This is an introduction to discourse analysis that presents both a theory of language-in-use and a method of research. Its aim is to demonstrate how language, both spoken and written, enacts social and cultural perspectives and identities.

Howarth, D. R., A. J. Norval, and Y. Stavrakakis (eds) (2000), *Discourse Theory and Political Analysis: Identities, Hegemonies and Social Change* (Manchester: Manchester University Press).
The case study chapters in this volume provide good examples of how discourse analysis can be applied.

Milliken, J. (1999), 'The Study of Discourse in International Relations: A Critique of Research and Methods', *European Journal of International Relations* 5(2) (June): 225–54.
This article addresses issues relating to research and methods in the use of discourse analysis, and how discourse should be studied, in International Relations.

Raymond, G. (2000), 'The Voice of Authority: The Local Accomplishment of Authoritative Discourse in Live News Broadcasts', *Discourse Studies* 2: 354–79.
This is a discussion of the authorized use of language. It focuses on the discursive practices that reporters use in live news broadcasts to exploit their right to authoritatively describe a scene all are witnessing, and then apply the findings of this exploration to wider discussions of this issue.

Ricento, T. (2003), 'The Discursive Construction of Americanism', *Discourse & Society* 14(5): 611–37.
This instructive case study of the construction of American national identity between 1914 and 1924 analyses texts from that period, which reveal competing discourses that are present today in American public discourse.

Shenhav, S. R. (2006), 'Political Narratives and Political Reality', *International Political Science Review* 27(3): 245–62.

Various authors (2004), Symposium: Discourse Analysis and Content Analysis, *Newsletter of the American Political Science Association Organized Section on Qualitative Methods* 2(1): 15–39.
This symposium discusses discourse and content analysis, separately and comparatively, and points to similarities as well as a number of important ways in which they differ.

Content analysis

Hopkins, D. and G. King (2010), 'A Method of Automated Nonparametric Content Analysis for Social Science', *American Journal of Political Science* 54(1): 229–47, available at: http://gking.harvard.edu/files/words.pdf
The authors present a method for overcoming problems encountered in coding blogs, speeches, government records, newspapers, or other sources of unstructured text.

Johnston, A. I. (1995), *Cultural Realism: Strategic Culture and Grand Strategy in Chinese History* (Princeton, NJ: Princeton University Press).
A highly effective example of content analysis using ancient Chinese texts to assess beliefs and predicted behaviour.

Kohlbacher, F. (2005), 'The Use of Qualitative Content Analysis in Case Study Research', *Forum: Qualitative Social Research* 7, available at: http://www.qualitative-research.net/index.php/fqs/article/viewArticle/75/153
A basic introduction to qualitative content analysis as an interpretation method for qualitative interviews and other data material. Useful for understanding how content analysis can contribute to qualitative case study research.

Krippendorff, K. and M. A. Bock (eds) (2008), *The Content Analysis Reader* (Thousand Oaks, CA: Sage Publications).
Fifty-one papers grouped around the following topics: the history and conception of content analysis, unitizing and sampling, inferences and analytic constructs, coders and coding, categories and data language, reliability, and validity, and computer-aided content analysis.

Lewis, R. B. (2004), 'NVivo 2.0 and Atlas.ti 5.0: A Comparative Review of Two Popular Qualitative Data-Analysis Programs', *Field Methods* 16(4): 439–69.
This article focuses on two popular qualitative data analysis tools that enable researchers to code text, sounds, pictures, or video, and to search codes for patterns, in order to help researchers decide which to use for the research tasks they have in mind.

Mayring, P. (2000), 'Qualitative Content Analysis', *Forum: Qualitative Social Research* 1, available at: http://www.qualitative-research.net/index.php/fqs/article/view/1089
The author describes an approach of systematic, rule-guided qualitative text analysis, including the central procedures of qualitative content analysis, inductive development of categories, and deductive application of categories.

Monroe, B. L., M. P. Colaresi, and K. M. Quinn (2008), 'Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict', *Political Analysis* 16: 372–403.
The authors discuss techniques that guide researchers, both in selecting words that capture partisan or other differences in political speech, and in evaluating their relative importance.

Neuendorf, K. A. (2002), *The Content Analysis Handbook* (Thousand Oaks, CA: Sage Publications).
Covers the history of content analysis, sampling message units, handling variables, reliability, and use of NEXIS for text acquisition. Also covers PRAM, software for reliability assessment with multiple coders. It also offers a very useful 'Flowchart for the Typical Process of Content Analysis Research'.

Riffe, D., S. Lacey, and F. G. Fico (2006), *Analyzing Media Messages: Using Quantitative Content Analysis in Research* (Mahwah, NJ: Lawrence Erlbaum).
A comprehensive guide to conducting quantitative content analysis. Provides step-by-step instruction on designing a content analysis study, and detailed discussion of measurement, sampling, reliability, data analysis, and validity.

References

Altheide, D. L. (1996), 'Reflections: Ethnographic Content Analysis', *Qualitative Sociology* 10(1): 65–77.

Austin, J. L. (1962), *How to do Things with Words* (Oxford: Clarendon Press).

Benoit, K., M. Laver, and S. Mikhaylov (2009), 'Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions', *American Journal of Political Science* 53(2) (April): 495–513.

Billig, M. and E. Schegloff (1999), 'Critical Discourse Analysis and Conversation Analysis: An Exchange between Michael Billig and Emanuel A. Schegloff', *Discourse & Society* 10(4): 543–82.

Brand, L. (2010), 'National Narratives and Migration: Discursive Strategies of Inclusion and Exclusion in Jordan and Lebanon', *International Migration Review* 44(1): 78–110.

Bryman, A. (2004), *Social Research Methods*, 2nd edition (Oxford: Oxford University Press).

Burnham, P., K. Gilland, W. Grant, and Z. Layton-Henry (2004), *Research Methods in Politics* (Basingstoke: Palgrave Macmillan).

Charteris-Black, J. (2006), 'Britain as Container: Immigration Metaphors in the 2005 Election Campaign', *Discourse and Society* 17(5): 563–81.

De Cillia, R., M. Reisigl, and R. Wodak (1999), 'The Discursive Construction of National Identities', *Discourse & Society* 10(2): 149–73.

Dixon, T. and D. Linz (2000), 'Race and the Misrepresentation of Victimization on Local Television News', *Communication Research* 27(5) (October): 547–73.

Downs, A. (1957), *An Economic Theory of Democracy* (New York: Harper).

Escobar, A. (1994), *Encountering Development: The Making and Unmaking of the Third World* (Princeton, NJ: Princeton University Press).
Fairclough, N. (1995), *Media Discourse* (London: Edward Arnold).
Fairclough, N. (2001), *Language and Power* (London: Longmans).
Fairclough, N. (2005a), 'Critical Discourse Analysis, Organizational Discourse, and Organizational Change', *Organization Studies* 26: 915–39.
Fairclough, N. (2005b), 'Blair's Contribution to Elaborating a New Doctrine of "International Community"', *Journal of Language and Politics* 4(1): 41–63.
Foucault, M. (1972), *The Archaeology of Knowledge and the Discourse on Language* (New York: Pantheon).
Frankel, C. (1969), *High on Foggy Bottom: An Outsider's Inside View of the Government* (New York: Harper and Row).
Hall, S. (1988), *The Hard Road to Renewal: Thatcherism and the Crisis of the Left* (London: Verso).
Hardy, C. (2001), 'Researching Organizational Discourse', *International Studies in Management and Organization* 31(3): 25–47.
Hardy, C., B. Harley, and N. Phillips (2004), 'Discourse Analysis and Content Analysis: Two Solitudes?', *Qualitative Methods: Newsletter of the American Political Science Association Organized Section on Qualitative Methods, Symposium: Discourse and Content Analysis* 2(1): 19–22.
Holliday, A. (2007), *Doing and Writing Qualitative Research*, 2nd edition (Thousand Oaks, CA: Sage Publications).
Holsti, O. R. (1969), C*ontent Analysis for the Social Sciences and Humanities* (Reading, MA: Addison-Wesley).
Hopf, T. (2004), 'Discourse and Content Analysis: Some Fundamental Incompatibilities', *Qualitative Methods: Newsletter of the American Political Science Association Organized Section on Qualitative Methods. Symposium: Discourse and Content Analysis* 2(1): 31–3.
Klingemann, H.-D., A. Volkens, J. Bara, I. Budge, and M. McDonald (2006), *Mapping Policy Preferences II: Estimates for Parties, Electors, and Governments in Eastern Europe, European Union and OECD 1990–2003* (Oxford: Oxford University Press).
Krippendorf, K. (2004), *Content Analysis: An Introduction to its Methodology*, 2nd edition (Thousand Oaks, CA: Sage Publications).
Laffey, M. and J. Weldes (2004), 'Methodological Reflections on Discourse Analysis', *Qualitative Methods: Newsletter of the American Political Science Association Organized Section on Qualitative Methods* 2(1): 28–30.
Laver, M., K. Benoit, and J. Garry (2003), 'Estimating the Policy Positions of Political Actors Using Words as Data', *American Political Science Review* 97(2): 311–31.
Lemke, J. L. (1998), 'Analysing Verbal Data: Principles, Methods and Problems', in B. J. Fraser and K. Tobin (eds), *International Handbook of Science Education* (Dordrecht: Kluwer), 1175–89.
Lowe, W. (2007), *Software for Content Analysis—A Review*, 22 March, available at: http://kb.ucla.edu/system/datas/5/original/content_analysis.pdf
Morgan, P. M. (1994), *Theories and Approaches to International Politics: What are we to Think?*, 4th edition (New Brunswick, NJ: Transaction Publishers).
Neuman, W. L. and K. Robson (2007), *Basics of Social Research: Qualitative and Quantitative Approaches* (Toronto: Pearson).
Potter, J. and M. Wetherell (1994), 'Analysing Discourse', in A. Bryman and R. Burgess (eds), *Analysing Qualitative Data* (London: Routledge), 47–66.
Proksch, S.-O. and J. B. Slapin (2009), 'Position Taking in European Parliament Speeches', *British Journal of Political Science* 40(3): 587–611.
Rapley, T. (2008), *Doing Conversation, Discourse, and Document Analysis* (London: Sage Publications).
Renner, J. and A. Spencer (2013), 'De-antagonising the Other: Changing Constructions of the Taliban and the Possibility of Reconciliation', *Global Society* 27(4): 475–96.
Rourke, L., T. Anderson, D. R. Garrison, and W. Archer (2001), 'Methodological Issues in the Content Analysis of Computer Conference Transcripts', *International Journal of Artificial Intelligence in Education* 12: 8–22.
Strauss, A. and J. Corbin (1990), *Basics of Qualitative Research: Grounded Theory Procedures and Techniques* (Newbury Park, CA: Sage Publications).
Titscher, S., M. Meyer, R. Wodak, and E. Vetter (2000), *Methods of Text and Discourse Analysis* (London: Sage Publications).
van Dijk, T. A. (2001), 'Critical Discourse Analysis', in D. Tannen, D. Schiffrin, and H. Hamilton (eds), *Handbook of Discourse Analysis* (Oxford: Blackwell), 352–71.
van Dijk, T. A. (2002), 'Discourse and Racism', in D. Goldberg and J. Solomos (eds), *The Blackwell Companion to Racial and Ethnic Studies* (Oxford: Blackwell), 145–59.
van Dijk, T. A. (2004), 'Racist Discourse', in E. Cashmore (ed.), *Routledge Encyclopedia of Race and Ethnic Studies* (London: Routledge), 351–5.
Weber, R. P. (1990), *Basic Content Analysis*, 2nd edition (Newbury Park, CA: Sage Publications).

Weldes, J. (1996), 'Constructing National Interests', *European Journal of International Relations* 2(3): 275–318.

Wesley, J. J. (2011), 'Observing the Political World: Quantitative and Qualitative Approaches', in K. Archer and L. Youngman-Berdahl (eds), *Explorations: A Navigator's Guide to Research in Canadian Political Science*, 2nd edition (Toronto: Oxford University Press).

Wetherell, M. (1998), 'Positioning and Interpretative Repertoires: Conversation Analysis and Post-structuralism in Dialogue', *Discourse & Society* 9(3): 387–412.

Wittgenstein, L. (1980) *Culture and Value* (Oxford: Blackwell).

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

Endnotes

1. The MRG was formed by Ian Budge and David Robertson, both at that time in the Department of Government, University of Essex. It was constituted formally as a Research Group of the European Consortium for Political Research (ECPR), which obtained funding to support most of its work in the 1980s.
2. The full coding process is described in Benoit, Laver, and Mikhaylov (2009).

Quantitative Analysis: Description and Inference

Chapter Summary

This chapter provides an introduction to quantitative analysis with a focus on description and inference. We discuss the different ways in which we can summarize data for a single variable from our sample and use it to make inferences about the wider population from which the sample was drawn. In particular, we focus on two of the key building blocks in quantitative research to do with measures of central tendency and measures of dispersion. The chapter contains:

- overview of analysis;
- univariate analysis;
- levels of measurement;
- methods of analysis;
- descriptive statistics;
- central tendency;
- dispersion;
- inferential statistics.

Introduction

Quantitative methods are among the most widely used techniques in political research. They are used to answer a wide variety of questions, to do with democracy and democratization (Przeworski et al. 2000; Gleditsch and Ward 2006), civil war (Gleditsch 2007) and ethnic conflict (Wilkinson 2004), and to do with public opinion (Wlezien 1995; Bartle, Dellepiane, and Stimson 2011), elections, and voting behaviour (Clarke et al. 2004, 2009; Franklin 2004). Whatever topic you are interested in studying in political research, it is hard to avoid coming across literature that uses quantitative methods. It is therefore an important method to understand. Yet for many students this is an off-putting prospect. To the untrained eye, quantitative research can seem intimidating and difficult to understand—closer to the study of statistics than politics. Yet in reality things are not as difficult as they

first appear. What can look complicated on the surface is actually relatively straightforward, even sensible, in practice, and familiarity with just a few basic concepts and techniques can take you a very long way in a short space of time. It is therefore well worth the effort of trying to learn how to use and understand quantitative methods, even if this involves some initial discomfort. It will not only help you understand the work of others, but, perhaps more rewardingly, it will also enable you to do your own original research. And these data-analysis skills are hard transferable skills that are also highly valued in the work place, and can help you get a job after you graduate.

In our experience students are sometimes apprehensive about learning quantitative methods. Yet in our experience this apprehension is soon overcome. Indeed, learning how to do quantitative research is often an incredibly rewarding experience. It is not only rewarding to overcome your initial anxieties, but in the process, new possibilities also open up, providing you with new skills which you can take with you in your own research and in your future career. Moreover, the actual statistical component of quantitative research is actually a relatively minor part of the research process. Good quantitative research is based upon good research design, interesting hypotheses, and a careful reading and understanding of the relevant literature. These research skills are common to all types of political research.

This chapter, and the chapters that follow, provide an introduction to the principles of quantitative research and a step-by-step guide on how to use and interpret a range of commonly used techniques. We start with the basics. The first part of the chapter looks at the building blocks of quantitative analysis. We focus on different ways in which data can be summarized, both graphically and with tables. We introduce two important measures: the **mean** and the **standard deviation**. In the second part of the chapter, we move on to inferential statistics, and discuss how we can make generalizations. In doing so, we introduce the concept of confidence intervals, often known as the margin of error.

Descriptive and inferential statistics

The research process involves the dual goals of description and explanation (see King, Keohane, and Verba 1994: chapter 2). Each is essential. We cannot construct meaningful explanations without a sound knowledge of what it is that needs to be explained. For example, there is a great deal of research which describes the political landscape. By describing the world around us, collecting data and information, we build up a picture of what is usual, or strange, or salient in some way, and what it is that is interesting to explain. Many of the questions we are interested in are therefore descriptive ones, the answers to which tell us something meaningful about the world we live in. Have people become less politically engaged? Is the media becoming more powerful? Are politicians corrupt? Is inequality increasing? How much power do domestic governments have over economic policy? What is globalization? Has it increased? These sorts of questions are all descriptive (see Chapter 4). Sometimes we are interested in describing the present, such as how much people know about politics, and sometimes we are interested in describing patterns over time, such as whether political trust is on the decline. Related to these descriptive questions, we then

develop explanations that we can seek to test. Why are some people more politically engaged than others? Why has trust in politics waned?

In the following chapters we will examine the statistical association between two variables (**bivariate analysis**) and three or more variables (**multivariate analysis**). This allows us to test hypotheses about causality, and answer questions about why things happen. But before getting into explaining why something happens, it is first a good idea to describe what happens in the first place.

Description in quantitative research comprises two parts. The first is to describe the data that we have collected. Whether it is from survey data (Chapter 11), expert interviews (Chapter 12), textual data (Chapter 14), or comparative aggregate data (Chapter 9), we can use descriptive statistics as a way of summarizing our sample data for a particular **variable**. But we can also use this information to try and make generalizations about the wider population from which the **sample** was drawn; that is, to make inferences. With inferential statistics, we can go from just talking about how many people in our sample support Brexit, to making inferences about what we think the true level of support for Brexit is in the UK population as a whole. Our ability to make this jump from describing our sample to making generalizations about the population rests on how we have collected the data in the first place. This is why probability samples are so important, because they create a statistical link between sample data that we have and the population that we want to make an inference about.

Summarizing the data

Next we turn to how different types of data can be summarized. There are a number of ways in which we can do this. We can use tables, figures, and statistics. We can look at the **frequency distribution**—which describes the entire distribution of responses, and summarizes the number of cases for each given response code. We can also summarize various aspects of this distribution, relating to measures of **central tendency**—such as averages—and measures of **dispersion** or variation—such as the standard deviation. The appropriate method for summarizing data depends upon the level of measurement (see Chapter 7). We discuss each of these in the next few sections.

Examining the distribution of variables also serves as a way of checking the data. One serious problem that affects quantitative analysis is **missing data**—or item non-response. For a variety of reasons, data for some cases in our data set may be unavailable. Respondents may not have answered some questions, or they may have answered but given ineligible responses. Missing data of this kind can introduce error or bias into the analysis if the answers of those people who did not answer the question would have been different in some important respect from those who did answer. It is not straightforward what to do about missing data, but there are a number of strategies that are available (see Box 15.1). The other thing to keep an eye out for is **sparse categories**, which are much easier to deal with. This occurs when relatively few observations are found for a particular response category. This can create problems when analysing the data, and so the best thing to do if this problem arises is to combine responses into larger, more general categories.

BOX 15.1 Missing Data

Missing data is a common problem in survey research (and other types of data collection). Running frequency tables allow you to identify the scope of the problem. Missing data can take different forms. Item non-response occurs when a respondent (for whatever reason) does not provide a valid answer to a question. It can reduce the sample size (especially in a multivariate analysis or when combining variables when several small amounts of missing data may accumulate) and risks introducing selection bias. Selection bias occurs when the sources of 'missingness' are not random, and are structured so that some types of people (or people holding some types of attitude or opinion) are less likely to provide a valid response than others. The structure of missing responses may vary from one variable to another, so there is no straightforward remedy that can be applied. The strategy for how to deal with this should depend in part on your theory about what has caused it. Below we discuss the common approaches to dealing with item non-response.

Listwise deletion is often the default choice, and simply involves deleting all missing values from the analysis. The danger with this approach is that you are throwing away information, reducing sample size, and more importantly, may be introducing bias if the missingness is not random. However, in cases where there are only a small number of missing values, this is probably the most practical solution. On the other hand, when there is a large number of missing values, it may be advisable to retain 'missing' as a separate category.

More sophisticated techniques involve trying to predict what respondents would have answered to the question from information that is available elsewhere in the questionnaire. This is known as *imputation*. The simplest form of imputation is just to set the missing values to the midpoint or mean. This approach is quick and easy, but makes best sense if you believe the missing value is random. In general, however, this is rarely the case, and setting missing to any constant value is regarded as unsophisticated and somewhat questionable. More advanced methods of imputation use regression methods (see Chapter 16) to impute values on the basis of other information provided by the respondent elsewhere in the questionnaire. Although this is statistically sounder, it can become quite technical, and is only useful in so far as good predictors are available.

Tables

Frequency tables are the normal tabular method for presenting distributions of a single variable. The table provides information on the distribution of responses across all response categories. They are therefore most appropriate when there are not too many different response categories (otherwise the tables become too big to easily interpret), and so are mainly used for nominal and ordinal variables. Tables should be clear and easy to understand. A good table presents the relevant information in a straightforward and transparent way. It should contain:

- a clear self-explanatory title;
- clear labels;
- an appropriate level of precision (round percentages to nearest whole number);
- sample size;
- a data source.

All tables should be clearly labelled and easy to understand. They should have a self-explanatory title, and be properly referenced with the source of data. The response codes

should be clearly labelled, and the sample size should be reported. If there is missing data, this needs to be reported too. If there is a large amount of missing data, it is important to be clear whether you are reporting percentages based on what is often called the total per cent (which includes missing values in the calculations) or the valid per cent (which does not include missing values).

For example, Table 15.1 summarizes the responses to the variable on vote choice that was asked in the 2017 British Election Study (BES) (Fieldhouse et al. 2017, 2018). It is a nominal variable. The first column contains a list of all the main response categories (the names of the main political parties in Britain). The second column (Frequency) contains information on the number of respondents from the survey who selected each response. At the bottom of this column in the last row, labelled 'Total', we can see that 2,194 people were interviewed for the survey. Of these people, 725 said they voted Labour, 671 said they voted Conservative, and 462 said they did not vote. We can use these numbers to calculate the percentage of the sample that provided each response. This is done in the next column, labelled 'Per cent'. The percentages are based on the total count. So, for example, 671 people out of 2,194 said they voted Conservative, and this corresponds to 30.6 per cent of our sample (671 divided by 2,194 multiplied by 100 = 30.6).

We should note, however, that some respondents did not provide an answer to the question. Some people did not vote, so if we are interested in the relative shares of the vote that each party received, we should exclude these people from the analysis. In addition some people did not know, or could not remember, which party they voted for, and for whatever reason, some people refused to answer the question. These 'missing values' are a common problem in survey research, and can pose a bit of a headache in terms of how we treat them. Do we regard them as valid answers, or do we discount them from our analysis?

Missing data can introduce error, since we cannot be sure what those people actually did. We can hope that they are a random selection of the sample. But it is also possible

Table 15.1 Voting behaviour in the 2017 British General Election

	Frequency	Percent	Valid percent
Labour	725	33.0	43.8
Conservative	671	30.6	40.5
Liberal Democrats	117	5.3	7.1
Scottish National Party (SNP)	49	2.2	3.0
Plaid Cymru	23	1.0	1.4
Green Party	31	1.4	1.9
United Kingdom Independence Party (UKIP)	37	1.7	2.2
Other	4	0.2	0.2
Did not vote	462	21.1	
Don't know	11	0.5	
Refused	64	2.9	
Total	2194	100	1657

Source: Fielding et al. (2017, 2018).

that they are more likely to be certain types of voter. For example, those that refused to answer the question may be more likely to be Conservative voters (the so-called 'shy Tories') or far-right voters (who may be unwilling to reveal their true behaviour because of social desirability bias; see Chapter 11), and so excluding them will bias our results, since it means that we are undercounting some response categories. When the number of missing cases is small (as in this case), this error or bias will also be small, but when the number is relatively large, it can be a problem and we need to think carefully about what to do about it.

The most straightforward option (though not always the most appropriate) is simply to discount the missing values from the analysis. We can then re-calculate the percentages to refer to just the people who provided valid responses. This is done in the next column, labelled valid per cent. Here the base excludes all those who said that they did not vote and all those who did not know who they voted for or who refused to answer. We can see that the percentages in this column are somewhat different. Now our base is just 1,657. So, for example, 671 out of 1,657 respondents said they voted Conservative, which corresponds to 40.5 per cent (671/1,657*100 = 40.5). This is slightly below the official vote share that the Conservatives received, which was 42.4 per cent indicating that for whatever reason the survey slightly under-reported the true level of support for the Conservatives and slightly over-reported the true level of support for Labour (we will come back to this point later in the chapter).

We should also note that we have very few responses for some of the minor parties. For example, less than 2 per cent reported voting for the Greens or Plaid Cymru. Sparse categories can pose a number of problems for analysis, and it is generally advisable to combine them where possible. In this instance, parties that got less than 5 per cent of the vote could be incorporated within the 'other party' category.

The next example we consider is from a question designed to measure whether respondents think of themselves as being ideologically left wing or right wing. The frequency distribution is presented in Table 15.2, and the exact survey question is reproduced under the table. The variable is ordinal. Again, we can summarize the data in terms of the frequency, the per cent, and the valid per cent, in the same way as before. However, we should note that there are quite a lot of people who selected the 'don't know' option. People may not have understood the question or may have been unable to place themselves. This in a sense is a valid answer. Not everyone thinks of themselves as being on the left or the right, and for analytical purposes we therefore need to think carefully about how we treat this data.

One option is simply to ignore the 'don't knows' and exclude them from the analysis. For descriptive purposes, we might be interested in the percentage of those who are able to place themselves on the left–right scale. This is reported in the column labelled 'valid per cent'. Here we have to be careful to be clear about what the data now refer to. They no longer refer to the ideological disposition of everyone (the British adult population), but only to those people who were able to place themselves. So, for example, of those who were able to place themselves on a left–right scale, 33 per cent placed themselves in the centre, and just 3 per cent placed themselves on the far left, and 2 per cent placed themselves on the far right. With ordinal variables we can also look at the cumulative per cent. The cumulative per cent is a

Table 15.2 Left–right ideology self-placement, British adults in 2017

	Frequency	Percent	Valid percent	Cumulative percent
0 Left	62	2.8	3.2	3.2
1	50	2.3	2.6	5.8
2	140	6.4	7.2	13.0
3	231	10.5	11.9	24.9
4	229	10.4	11.8	36.6
5	639	29.1	32.9	69.5
6	178	8.1	9.2	78.7
7	222	10.1	11.4	90.1
8	128	5.8	6.6	96.7
9	20	0.9	1.0	97.7
10 Right	44	2.0	2.3	100.0
Don't know	239	10.9		
Refused	12	0.5		
Total	2194	100	1943	

Note: Question: 'In politics, people sometimes talk about parties and politicians as being on the left or right. Using the 0 to 10 scale on this card, where the end marked 0 means left and the end marked 10 means right, where would you place yourself on this scale?'

Source: Fielding et al. (2017, 2018).

useful way of aggregating different responses. It adds up the percentages in order for each response category. From this, we can clearly see that nearly 37 per cent of respondents who were able to place themselves on a left–right scale placed themselves to the left of centre (giving answers 0 to 4).

With interval variables it is generally not feasible to report the data in tables, since there will be a lot of different response categories, and so the table will require a lot of rows. A table of age, for example, in which respondents' age varies from 18 to 100 would require eighty-two rows, which would go on for several pages. There would also be a lot of sparse categories. The purpose of tables is to present data in a clear and transparent way. Too many rows can distract from this. Moreover, it would not be very easy to interpret. One option then is to recode data into bands. So rather than displaying the distribution for all ages, we can band the ages into age groups (e.g. 18–30 years old, 31–40 years old, 41–50 years old, 51–64 years old, and 65 years and over). There are a number of different ways in which data like these can be grouped together. The data can be grouped into equal percentiles, so that each band contains, say 20 per cent of the sample; they can be grouped into equal intervals, so that each band spans, say, fifteen years; or they can be grouped into distinctive categories of theoretical interest, so that bands refer to specific groups of interest, such as young adults or retired people. It doesn't really matter which approach is used, but it is advisable to avoid having groups with either a very small or a very large sample size.

Graphs and figures

Sample distributions can also be displayed graphically. This is often a more accessible way of presenting data, and can be easier to analyse. The purpose of graphs and figures is to present the data as clearly and accurately as possible. People can often get very creative when it comes to displaying graphs, but it is wise not to get carried away, and to focus on just trying to present the information in the clearest terms possible. In most cases, a **bar chart** or **histogram** is the best option for summarizing the distribution of a single variable. **Line graphs** are also frequently used for summarizing data over time. However, the alternatives, such as pie charts, generally are not to be recommended. They may look pretty, but it can often be difficult to distinguish between the relative sizes of the different 'slices'. However, this is not such a problem with histograms and bar charts, which generally present the data more clearly.

The choice between bar charts and histograms depends upon the level of measurement of the variable. For nominal and ordinal variables bar charts are used, and for interval variables histograms are used. In practice, bar charts and histograms look quite similar to each other. The principal difference is that bar charts have a space between the bars to indicate that the response categories are distinct—or **discrete**—whereas with histograms the bars all touch each other to indicate that the response categories are **continuous**.

Figure 15.1 presents the data from Table 15.1 as a bar chart. Along the x-axis (the horizontal axis), the different response categories are labelled. The y-axis (the vertical axis) records the reported share of the vote in per cent. It is important to clearly label the axis so that the reader can interpret the information. Presenting the data in this way conveys all the relevant information, and clearly shows which parties received the most votes and which the fewest.

Figure 15.2 depicts a bar chart for left–right self-placement. This variable is ordinal, and so there are spaces between the bars, although 0–10 scales are often treated as interval variables and so it could be presented as a histogram. Presenting the data in this way is in some

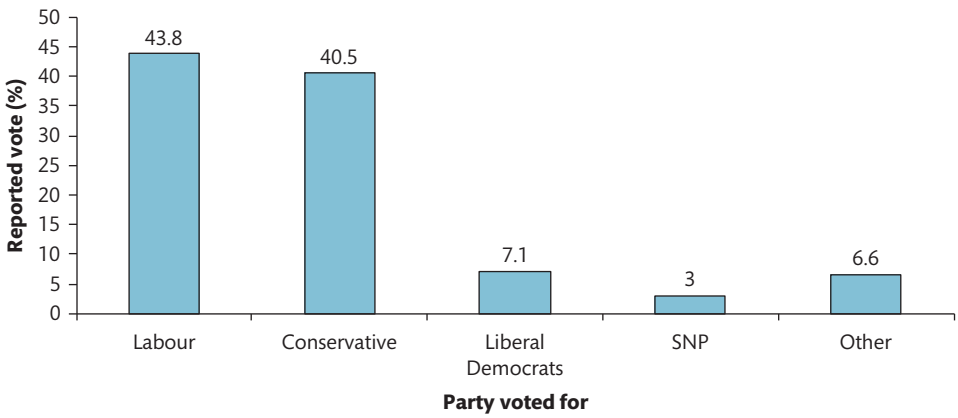


Figure 15.1 Bar chart of reported party vote in the 2017 British General Election

Source: Data from British Election Study 2017. https://www.britishelectionstudy.com/news-category/2017-general-election/.

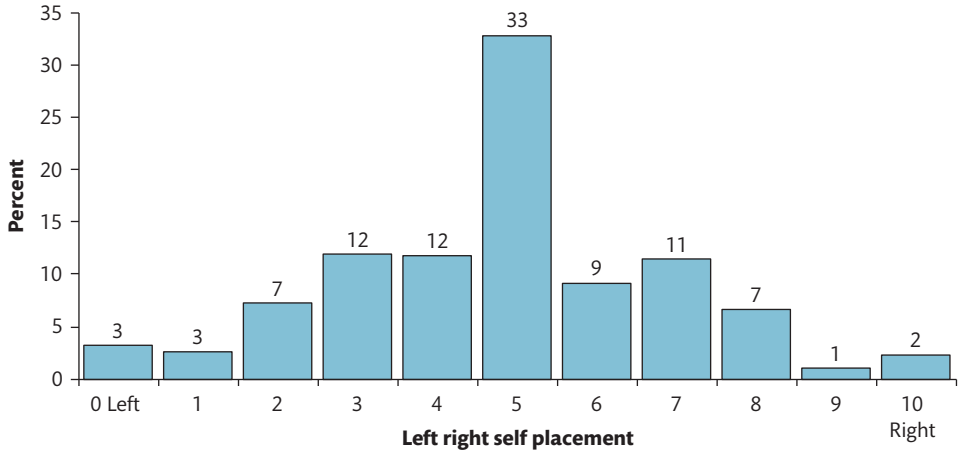


Figure 15.2 Bar chart of left–right self-placement

Source: Data from British Election Study 2017. https://www.britishelectionstudy.com/news-category/2017-general-election/.

respects an improvement on just presenting the data in tabular form, since we now get a clearer sense of the overall distribution of the variable. We can see that the most common responses are in the centre of the distribution, and the least common responses are in the **tails** of the distribution (the extreme values at either end). In fact, the variable looks quite like a **normal distribution**, which is something we will be coming back to later on.

Figure 15.3 depicts a histogram for the Scottish National Party's (SNP) share of the vote in the constituencies which it contested for the 2017 British General Election. This is an example of **aggregate data**. Since share of the vote is an interval-level variable, we display the data in a histogram, where the bars touch each other to indicate that the data are continuous.

As we discussed with reference to Table 15.1, very few survey respondents reported having voted for the SNP in the 2017 election. The party secured just under one million votes, which across the whole of the UK equates to just 3 per cent of the national vote. The BES survey data is therefore not well-placed to analyse the factors associated with the SNP vote, since there is not enough information to draw sensible conclusions. An alternative approach then is to examine the official election returns, and see how well the party did in each constituency which it contested. This is reported in Figure 15.3. The SNP unsurprisingly only contested in Scotland, where it stood in all fifty-nine constituencies. In these constituencies its share of the vote ranged from a low of 22 per cent to a high of 47 per cent. Presenting the data in this way we can get a sense of how the level of support for the SNP varies across constituencies. Even though across the UK as a whole it only got 3 per cent of the national vote, in those constituencies in Scotland where it contested it received 37 per cent of the vote. From Figure 15.3 we can get a rough idea of how well it did on average (about 40 per cent appears to be the most common value) and a rough idea of how much its vote varied, but to make clearer statements about these types of observation we need to employ special statistics. These are discussed in the next section.

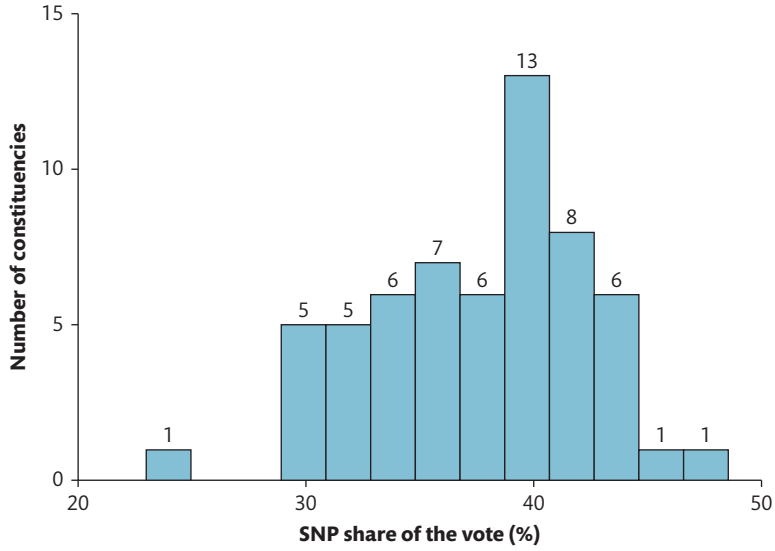


Figure 15.3 Histogram of Scottish National Party share of the vote in constituencies in the 2017 British General Election

Source: Fielding et al. (2017).

Summarizing the distribution using statistics

As well as summarizing the entire distribution, we can also choose to summarize some specific aspect of the distribution. This allows us to more easily make comparisons, either over time or between groups of people, countries, or places. In making these comparisons, it is useful to have information on the centre of the distribution, and how scattered or dispersed the data are around this point. These measures are known as measures of central tendency and measures of dispersion.

Measures of central tendency

There are three measures of central tendency: the mode, the median, and the mean. Each measure provides a summary of the 'central value' of the distribution in some sense, although how they do this varies somewhat. These measures are generally only appropriate for ordinal and interval-level data, since for nominal-level variables it does not really make much sense to talk about the central value because the values do not have any underlying order to them. However, for nominal variables, we are often interested in the mode, which refers to the most common value, or the largest group. So we might say that the modal religion in India is Hinduism, or the modal ethnic group in Kenya is Kikuyu, or the modal political party in the UK parliament is the Conservative Party.

To illustrate how these measures of central tendency are calculated, consider the following example. Suppose we were interested in the financial status of students, and decide to ask each of the students in a class how much money they receive from all sources per year.

There are eleven students in the class, and each is asked to estimate their total annual income in pounds (to the nearest thousand). From this we obtain the following values. This is our distribution:

4, 6, 8, 8, 10, 10, 10, 12, 14, 16, 20

The **mode** is simply the most common value in the distribution. We can see that three students have an income of about £10,000. This is the most common value in the distribution and so is the mode. It is possible to have more than one mode in a distribution; in such cases, the distribution is bimodal or multimodal. The **median** refers to the middle value of the distribution. It is defined as the point which divides the observed values into two equal parts. All the values are arranged in ascending order (as above) and the median is simply the one in the middle. Since we have eleven students, in this case the median refers to the value of the sixth observation, which is also £10,000. If there is an even number of observations in the distribution (and so there is not one value that is directly in the middle), the median is calculated as the difference between the two middle values.

Whereas the mode and the median can be directly observed from the distribution, the mean needs to be calculated (though most computer programs will do this for you). The mean is the measure of central tendency that most of us think about when we hear the term 'average'. It is symbolized in two different ways, according to whether we are describing the population mean or the sample mean. For the time being, we just consider the sample mean. This is denoted as $\overline{X}$ (pronounced x-bar), where x is the variable of interest (in this case, student income). The mean is the measure of central tendency that is used most often in quantitative analysis, and is one of the building blocks for more advanced quantitative analysis techniques that we will encounter in the following chapters. It is calculated by adding up all the values of x that are observed in the distribution, and then dividing the total by the number of observations (students). So, in the above example, the mean is £10,720 (see Box 15.2).

When we have a perfect normal distribution the mode, the median, and the mean are all the same. However, it is not uncommon to find that sometimes the different measures produce very different values. Under these circumstances, it is natural to ask which measure is the best, and it is easy to suppose that it must be the mean, since this appears to be the most scientific. But this is not always the case. The mean can be sensitive to extreme outliers, especially when we are dealing with relatively small samples. Suppose the day after we collected the above data, a new student entered the class who had a trust fund which gave them an annual income of £2 million. Including this person in our calculation dramatically changes the mean. Whereas the mode and the median are still £10,000, the mean is now a whopping £166,676. Clearly, the mean no longer appears to be very typical of what the average student has to live on. It is therefore always important to check for the presence of outliers (particularly when the sample size is relatively small), as these can dramatically distort the results and make what is highly unusual appear typical.

But at the same time, the problem with the median is that it is quite a blunt measurement instrument. If we are interested, for example, at looking at how public opinion has changed over time, we might prefer a more sensitive measurement instrument. For example, Eurobarometer routinely asks people in different countries around Europe to place themselves on a

BOX 15.2 How to Calculate the Mean

The formula for calculating the mean is expressed as follows:

$$\bar{X} = \frac{\sum X_i}{n}$$

where

$\bar{X}$ = the mean;
$\sum$ = 'the sum of' (pronounced sigma);
x_i = the values of x to be summed;
n = the number of observations (sample size).

Using the above example, we can plug in all the values of x, add them up, and then divide by the number of observations (11). This gives us:

$$\bar{X} = \frac{4+6+8+8+10+10+10+12+14+16+20}{11}$$

$$\bar{X} = \frac{118}{11} = 10.72$$

0–10 left–right scale, where 0 is left and 10 is right (as discussed earlier). Figure 15.4 plots the median position for Europeans since 1973 using data from Bischof and Wagner (2019). We can see that the median left–right position in Europe has not changed since the 1970s and has remained constant at 5, suggesting that there has not been any change in public attitudes.

By contrast, Figure 15.5 plots the mean position since 1973. From this measure we can see that in the 1980s and 1990s there was a gradual shift to the left in European public opinion. At the turn of the millennium public opinion turned to the right. But since the financial crisis in 2008 there has once again been a leftward turn. Granted, none of these

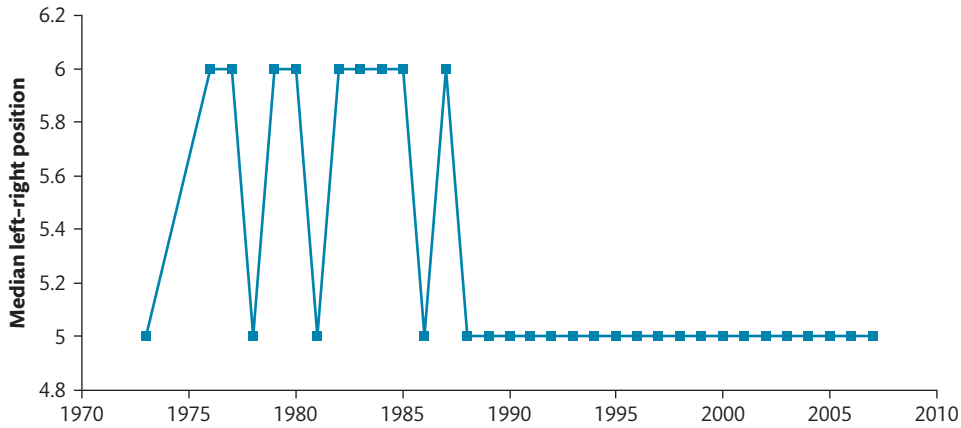


Figure 15.4 Left–right self-placement in Europe—median position of adults, 1973–2016

Note: Question: `In politics, people sometimes talk about parties and politicians as being on the left or right. Using the 0 to 10 scale on this card, where the end marked 0 means left and the end marked 10 means right, where would you place yourself on this scale?'

Source: Eurobarometer.

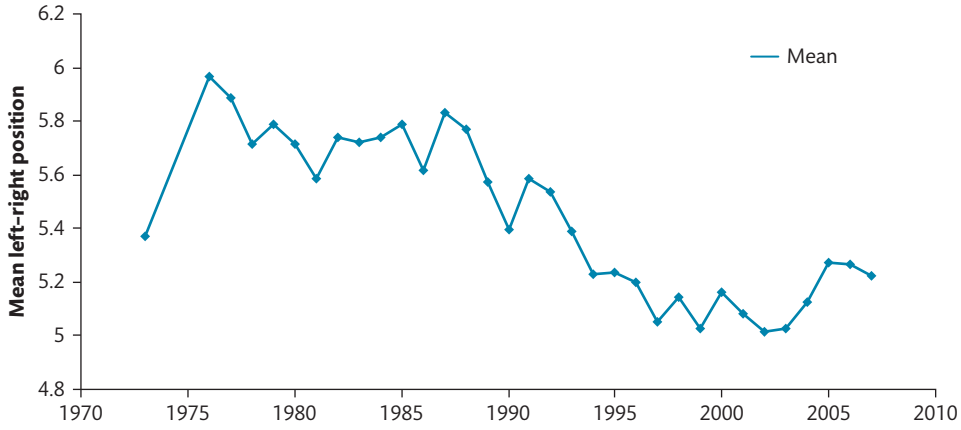


Figure 15.5 Left–right self-placement in Europe—mean position of adults, 1973–2016

Source: Eurobarometer.

changes have been particularly dramatic—and public opinion still generally occupies the centre ground, but even small changes at the aggregate level can have important electoral implications (see Bischof and Wagner 2017, 2019).

Choosing which measure of central tendency to report depends upon the kind of information you want to convey. The mode is appropriate if you are talking about data measured at the nominal level, or if you are simply interested in what is the most common value or the largest group. The median is appropriate if there are extreme outliers, which may distort the mean. Most government publications use the median rather than the mean to report average household income for just this reason. Otherwise, the mean is generally appropriate. The mean is the only measure that is mathematically based. It uses all of the information in the distribution, and is one of the main building blocks for more advanced analysis.

Measures of dispersion

Measures of central tendency provide a powerful way of summarizing large amounts of data. They are useful for summarizing what is the average or typical person or value. But it is also useful to get an idea of how individuals or objects in the distribution differ from one another. Sometimes two different distributions can have the same mean (or median) but be very different from one another in terms of the variability or spread within the distribution. An example illustrates this. Suppose students taking introduction to quantitative methods are assigned to one of three classes. Each class has twelve students, and at the end of the course the students are given a test that is marked out of 10. The marks for each student in each class are:

Mark out of 10

Class 1: 6 6 6 6 6 6 6 6 6 6 6 6

Class 2: 4 4 5 5 5 6 6 7 7 7 8 8

Class 3: 1 2 2 3 4 4 8 9 9 10 10 10

For each class, the measures of central tendency based on the mean and the median are the same—6—which represents a pass mark. However, the variation in the marks is different in each of the classes. In Class 1, everyone gets the same mark, perhaps because they have just been taught the bare minimum to pass, and this has been drummed into them over and over again during the course. In Class 2, there is more variation in the marks—some students do well whereas others fail, but there is not much difference between the top marks that are achieved and the bottom marks. In Class 3, there is more variation still. The students tend to do either very well or very badly, perhaps because the teacher focuses more on the able students and leaves behind the students who are struggling.

This example shows how important it is to examine both the variation and the mean. On the face of it, we might assume that there is not much difference between each of the classes since they all have the same mean and median exam mark. However, from inspecting the variation of marks within each class, we might come to the conclusion that the classes are in fact quite different from one another. And, depending upon how confident we felt in terms of our own ability to do quantitative research, we might have strong opinions about which class we would rather go into.

This example clearly illustrates the intuition behind calculating measures of dispersion. As with the measures of central tendency, variation—or dispersion—can be measured in a number of ways that vary in statistical complexity. The three most commonly used measures of dispersion are: the range; the **interquartile range** (IQR), and the standard deviation. These are discussed below.

The **range** is the most straightforward measure of dispersion. It simply calculates the difference between the smallest and the largest values. So with reference to the above example, the range for Class 1 is zero (6 – 6 = 0), the range for Class 2 is 4 (8 – 4 = 4), and the range for Class 3 is 9 (10 – 1 = 9). Even this simple measure gives us a better feel for the data than we would have from considering just the mean alone. We know that there is a lot of variability in the marks of students in Class 3, and none at all in Class 1. So if we were a student who was particularly apprehensive about taking a quantitative methods course, we might feel more confident if we had the teacher for Class 1.

However, one drawback of the range is that, since it depends on the extreme values of the distribution, it is very sensitive to outliers. One or both of these values might be atypical, so the range is not always a particularly trustworthy measure. The IQR avoids this problem by ignoring the tails of the distribution, and only examines the dispersion within the middle 50 per cent of the distribution. It thus divides the distribution into quartiles and compares the value of the first quartile (the value which divides the ordered set of observations into the smallest 25 per cent and the largest 75 per cent) and the third quartile (the value which divides the ordered set of observations into the smallest 75 per cent and the largest 25 per cent). This can be thought of as a related technique to the median, which divides the data into the smallest 50 per cent and the largest 50 per cent.

Table 15.3 provides a worked example with data from Class 3. We can see that all the students in the bottom quartile got less than 3 out of 10 on their test, and that all the students in the top quartile got full marks of 10 out of 10. The IQR (the middle 50 per cent) therefore stretches from marks of 3 out of 10 to 9 out of 10, giving a range of 6.

The standard deviation is related to the mean, and is by far the most widely used measure of dispersion. Rather than just considering the difference between two points of the

Table 15.3 Calculating the interquartile range

Student	X_i (Test score)	
1	1	
2	2	Lower quartile
3	2	
4	3	
5	4	
6	4	Interquartile range
7	8	
8	9	
9	9	
10	10	
11	10	Upper quartile
12	10	

distribution (as with the range and IQR), it utilizes all the values in the distribution and describes how far on average each value deviates from the mean. When all the values in the distribution are the same (as in the example from Class 1), the difference between each value and the mean will be zero. However, when the values are very different from the mean (as in Class 3), the standard deviation will be relatively large. The formula for how to calculate the standard deviation is shown in Box 15.3.

The standard deviation is the square root of the sum of the squared differences between each value in the distribution and the mean, divided by the sample size minus one. This sounds a bit of a mouthful, but is actually fairly straightforward if we consider each step in turn. Table 15.4 provides a worked example for calculating the standard deviation, using the distribution of test scores from Class 3, just discussed. We have twelve students in our sample, and for each student we have recorded their test score. These test scores are the values for x_i, and are listed in column 2 for each student. For example, we can see that student 1 got just 1 out of 10 on the test. We are interested in how far these scores deviate from

BOX 15.3 How to Calculate the Standard Deviation

The standard deviation is calculated with the following formula:

$$s = \sqrt{\frac{\Sigma\left(X_i - \bar{X}\right)^2}{N-1}}$$

where

s = the standard deviation of the sample;
x_i = the value for each observation in the distribution;
$\bar{X}$ = the sample mean;
N = the number of observations (sample size).

Table 15.4 Calculating the standard deviation

Student	X_i (Test score)	$(X_1 - \bar{X})$	$(X_1 - \bar{X})^2$
1	1	-5	25
2	2	-4	16
3	2	-4	16
4	3	-3	9
5	4	-2	4
6	4	-2	4
7	8	2	4
8	9	3	9
9	9	3	9
10	10	4	16
11	10	4	16
12	10	4	16
Sum (Total)			144
Sum/*N*–1			144/11 = 13.09
Square root			3.62

the mean. We already know that the mean is 6, so for each student we calculate how far away their score is from 6. This is done by subtracting the mean from each individual score. This is shown in column 3.

For student 1 this comes out as –5 (1–6 = –5). The problem is that some of these deviations are positive and some are negative, so if we summed them all up, the positives and negatives would cancel each other out and we would be left with zero. To get rid of the different signs, we therefore square the deviations. This is presented in column 4. Next we add up these values. This comes to 144. This value is known as the sum of the squares. We then divide this value by our sample size minus one (11) which gives us 13.09. This value is known as the **variance**. The last step involves taking the square root of this number, which gives us the standard deviation. This is 3.62. Some students got scores which were close to the mean, other students got scores which were far away from the mean, but what the standard deviation tells us is that, on average, the deviation from the mean was 3.62 points.

Like the mean, the standard deviation is one of the cornerstones of quantitative analysis. It can be used to summarize the dispersion of a single variable within our sample. It can also be used to compare the dispersion among two or more samples. This is often of more analytical interest. It can be hard to know how to interpret the size of a standard deviation when there is only one value. Is it big or is it small? But when we have repeated values, we can look at whether the standard deviation changes, and if so, whether it gets bigger or smaller. This is often of considerable analytical interest. Measures of variation have many practical applications in political research, and can be used to tell us about extremism, polarization, and inequality. For example, suppose we are interested in ideological extremism. There is a great deal of talk these days about ideological polarization, meaning that voters have vacated the centre ground and moved towards holding more extreme postions.

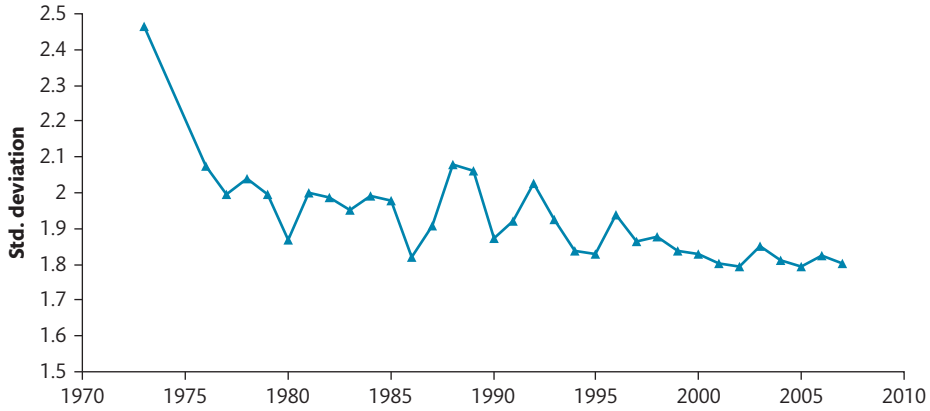


Figure 15.6 Left–right self-placement in Europe, standard deviation, 1973–2016

Source: Eurobarometer.

One way of examining this is just to look at the mean, as we did earlier in Figure 15.5. From this figure, we can see that the centre has shifted, but that it appears to shift in a cyclical fashion. However, it doesn't tell us whether voters have converged or become more polarized. Another way of examining this then is to look at the standard deviation of left–right scores. This tells us how clustered or spread out the scores are. Figure 15.6 plots the standard deviation of left–right self-placement among adults in Europe. There appear to have been some noteable changes over time. In the earlier period in the 1980s and 1990s the standard deviation of left–right values decreased, indicating that people were becoming less polarized on left–right issues and converging on the centre ground. However, since the turn of the millennium the standard deviation has increased, indicating a period of heightened polarization (see Research in Action Box 15.1).

Research in Action 15.1 Do Voters Polarize When Radical Parties Enter Parliament?

In recent years academics have paid a great deal of attention to the rise of radical and extreme parties in Europe and elsewhere and what impact if any they have on the functioning of democracy. One aspect of this is to do with what happens when they enter parliuament. Does the election of extremists normalize extremist ideology? Or does it produce a public backlash? In the wake of Brexit and the election of Donald Trump, these questions have become even more salient as a number of countries appears to be experiencing a period of heightened political antagonisms and increased polarization. To explore these issues Daniel Bischof and Markus Wagner (2019) investigated the impact of extreme parties on public opinion when those parties are elected to parliament. To do this they analyse patterns of left–right polarization after radical-right party entry in seventeen countries from 1973 to 2016 based on Eurobarometer data. They find evidence that the public polarizes after extreme party entrance. Moreover, the size of this effect is substantial, at about half of a standard deviation. The results of the study indicate that first-time radical-right party entry leads to voter polarization in both the short term and the long term. This is because the legitimation of extremists and the backlash against extremists occurs simulatensously, meaning that voters on both sides of the ideological spectrum move to the extremes.

Descriptive inference

So far we have considered ways in which we can describe our sample. But we are also interested in making inferences about the population from which the sample was drawn. This allows us to make generalizations, and it is this ability to make robust generalizations about the wider population that is one of the key strengths of quantitative analysis. In order to make generalizations, we use the data from our sample to tell us something—or make inferences—about the wider world. Inferential statistics allow us to build upon what we have covered so far and to make estimates about population values. These estimates often take the form of a **confidence interval**, which is a range of values within which we are fairly sure the 'real' population value is contained. This idea of the confidence interval is particularly useful when we are interested in comparing values over time or between countries, and helps us to decide whether changes or differences are 'meaningful' or not.

Before outlining how inferential statistics work in practice, it is worth discussing a little of the theory behind the approach. We do not go into great detail here, since our focus is on the analysis and interpretation of these statistics rather than how they work, but useful summaries can be found in Agresti and Finlay (2009). Our sample is interesting. We hope it is reliable, and we hope that it is representative so that we can make inferences about the population from which it was drawn. But how exactly can we do this? The **central limit theorem** is what allows us to link our sample to the population. The central limit theorem is perhaps the most important theorem in statistical theory. It states that if repeated random samples are drawn from a population, the **sampling distribution** of the sample estimate (e.g. Conservative vote share) will approach normality. An example helps to illustrate this.

Suppose we interview a hundred people and find that forty-five said they would vote Conservative if there were an election tomorrow. We can think of this as our sample estimate. We hope that our sample estimate is a good approximation of the true population value, and that in the real world somewhere around 45 per cent of the population would indeed vote Conservative. But if we interviewed a different hundred people, we might find that only forty-four people said they would vote Conservative, and if we interviewed a different hundred people again, we might find that only forty people said they would vote Conservative, and so on. This might sound like a potential problem, but the beauty of the central limit theorem is that it isn't. We know that each of our samples is likely to produce a slightly different estimate, and we also know that if we take a large number of random samples (of sufficient sample size), most will come up with a percentage estimate close to the true population value. In only a few samples will the sample estimate be way off the mark. In fact, the distribution of the sample estimates (known as the sampling distribution) would approximate a 'normal' distribution (see Figure 15.7).

We can think of the sampling distribution as being a distribution of all the possible sample estimates that we could draw from a population using the same sample size. If we were to take all these possible sample estimates, sum them, and divide them by the number of samples, we would have the exact value of the population. The elegance of the sampling distribution is that even if the population is not normally distributed, repeated samples will generate estimates that are approximately normally distributed. And, as with

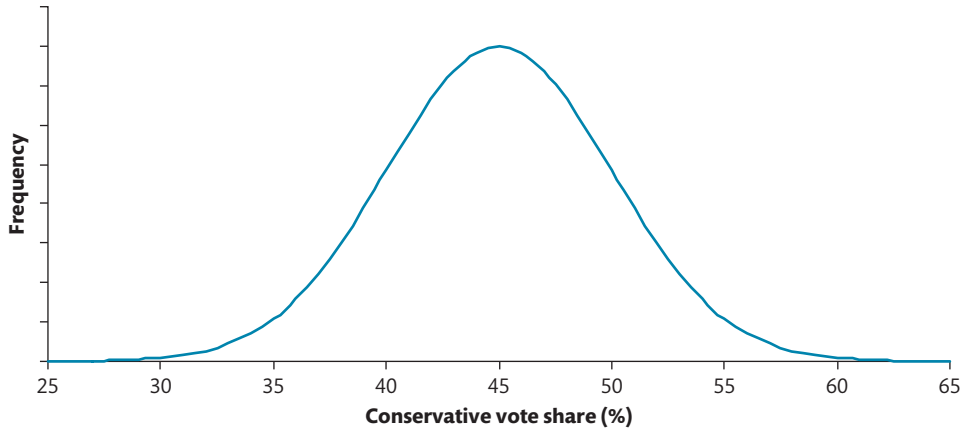


Figure 15.7 The normal distribution

all distributions, the sampling distribution will have a mean, and this mean will be the true population value. The peak of the sampling distribution therefore corresponds to the true population value for Conservative share of the vote (45 per cent).

Now, we hope that our sample estimate is exactly the same as the population value or at least pretty close to it, but in truth we do not know where it is. Our sample estimate could potentially be one of the extreme estimates right off in the tails of the distribution, a long way away from the true value of the population. However, if it was, we could console ourselves that we must have been pretty unlucky to have drawn such a duff sample if we had done everything right in the sampling phase of the research (see Chapter 11). Crucially, it is this idea of how lucky or not we would have to be to draw a duff sample that underpins inferential statistics. So how do we decide whether the sample we have drawn is one of the samples that is close to the true population value or one that is far away?

To answer this question we need to know about something called the **standard error**. This can be used to tell us how likely it is that our sample estimate is close to the true population value. As with all distributions, the sampling distribution has a standard deviation. Whereas the mean of the sampling distribution is the same as the population value, the standard deviation of the sampling distribution is smaller than the population standard deviation. This is because there tends to be more variation between values for individuals than there does between statistics summarized for samples. This makes intuitive sense. According to the Office of National Statistics the median income for full-time employees in Britain in 2015 is around £27,000 per year. Individuals earning £50,000 per year or more are uncommon, though not unheard of (just over 10 per cent). However, a random sample of 1,000 people with a median income of £50,000 per year would be exceptionally rare.

The standard deviation of the sampling distribution has a special name: the standard error. Probability theory tells us that in 95 per cent of samples that we draw, the population value will be within two (actually 1.96) standard errors of the sample estimate. With nominal variables, we can estimate the standard error of a proportion. With interval (and

at a stretch ordinal) variables we can estimate the standard error of a mean. The way in which we do this is slightly different, but in both cases the size of the standard error is partly a function of sample size. These points are not only of statistical interest, but are of great practical use as well. We can utilize the idea of the sampling distribution and the standard error to make inferences about populations. This is discussed in the next section.

Population estimates

As soon as we move into the realm of inferential statistics, there are a whole host of things we can do. The first, most basic application is that we can make guesses about the true population value that we are interested in. We can do this for nominal, ordinal, and interval data. For nominal variables, we can make a guess about the population proportion, such as what is the proportion of the population that will vote Conservative. For ordinal and interval variables, we can also make a guess about the population mean, such as what is the mean position on the left–right scale of the population as a whole.

Confidence intervals for proportions

One common use of surveys and opinion polls is to track likely voting intentions in the general election. This is something we frequently read about in the newspapers. For example, a survey may report that if there were a general election tomorrow, 44 per cent of the sample say they would vote Conservative. This is interesting to know, but more interesting is what it tells us about our estimate of what the true level of support for the Conservative party is in the population as a whole. We think it is probably around 44 per cent, but we would be unwise to state that it is exactly 44 per cent. What we have is an estimate.

It is common to accompany this estimate with a margin of error. Survey companies often say that the margin of error is plus or minus three percentage points. This means that the best guess is 44 per cent, but the expected true value could be between 41 and 47 per cent (44 per cent plus or minus three percentage points). This is quite a big range. During election campaigns, commentators can get quite excited about small changes in the polls, but most of the time, when a party's share of the vote appears to go up or down by a few percentage points, it is still well within the margin of error. It does not mean that the underlying level of support for the party in the population has changed, but rather it reflects what is known as sampling variation.

The technical term for this margin of error is a confidence interval. This section describes the calculation behind a confidence interval, though in practice it is not something that you need to know how to calculate, since most statistical software programmes will do it for you. To calculate a confidence interval we draw on probability theory, and the ideas of the sampling distribution and the standard error discussed earlier. Since we know that 95 per cent of the time the population value will fall within 1.96 standard errors of the sample estimate, we can calculate the range within which we think the true population value lies. To estimate this we need to calculate the standard error. Because we are dealing with a

nominal variable, we calculate the standard error for a proportion. But the procedure is much the same as that for calculating the standard error for the mean. See Box 15.4.

Having calculated that standard error, we can now estimate the range within which the population value is likely to be. This is our confidence interval (see Box 15.5). If we specify a 95 per cent confidence interval, it means that we think our range will correctly contain the true population value 95 per cent of the time. So in percentage terms we can say that we think the true level of Conservative support lies somewhere between 41 per cent and 47 per cent (44 per cent plus or minus three percentage points). This is our best guess. Of course, we don't know whether our guess is correct or not. We may be wrong. But we can say that most of the time our guess will be correct. In fact, we can say that 95 per cent of the time our guess will be right. This, of course, leaves us with the possibility of being wrong 5 per cent of the time, which will happen when we draw one of those dodgy samples right off in the tails of the sampling distribution. This is known as sampling error, which is the culprit frequently blamed when surprising one-off polls are recorded which seem out of step with what else has been recorded.

There is thus a trade-off between precision and confidence. If we were not happy to be correct 95 per cent of the time, we could extend our confidence interval. Probability theory tells us that 99 per cent of the time the population value will be within 2.58 standard errors of the mean. This will give us a confidence interval for the true level of Conservative support of between 40 per cent and 48 per cent (44 per cent plus or minus four percentage

BOX 15.4 How to Calculate the Standard Error for a Proportion

The formula for the standard error of a proportion is:

$$SE_p = \sqrt{\frac{(P_s)(1-P_s)}{N}}$$

where

SE_p = the standard error of the proportion;
P_s = the sample proportion;
N = the sample size.

BOX 15.5 How to Calculate the Confidence Interval for a Proportion

This formula for the confidence interval of a proportion is expressed as:

$$ci = P_s \pm 1.96\sqrt{\frac{(P_s)(1-P_s)}{N}}$$

If we assume our sample size is 1,000, and the percentage who say they will vote Conservative is 44 per cent, then we can plug in the numbers as follows:

$$ci = 0.44 \pm 1.96\sqrt{\frac{(0.44)(0.56)}{1000}}$$

$$ci = 0.44 \pm 0.03$$

points). Now we have more confidence that our estimate will contain the true population value, but our estimate is wider and less precise. For this reason, the 95 per cent confidence interval is generally accepted as the default position.

Having established a confidence interval for one proportion, we can then use this to make a comparison with the confidence interval of another proportion, to see if the difference, say between the Labour and Conservative share of the vote, could just be down to sampling variation or whether it represents a real 'statistical difference'. So, for example, if just before an election we carry out a survey and find that the Conservatives are on 44 per cent and Labour are on 41 per cent, how confident can we be that this represents a 'real' difference in public opinion?

Our sample estimate for Labour may be 41 per cent, which is somewhat lower than our estimate of the Conservative vote share, but before rushing to any conclusion, the important thing is to inspect the confidence interval for Labour's vote share. The 95 per cent confidence interval for Labour vote is between 38 per cent and 44 per cent (41 per cent plus or minus three percentage points). There is thus some overlap between our two estimates. It could be that the true population value is at the top end of the Labour estimate and the bottom end of the Conservative estimate, in which case there may be no real difference in their respective shares of the vote.

This leads us to a slightly different question that we want to answer. Rather than trying to find out what the true level of support for each party is in the population, we now want to know whether there is a real difference between their vote shares. Do we have enough evidence to reject the idea that they have equal support? This introduces us to the idea of hypothesis testing and the null hypothesis (which we examine in more detail in Chapter 16). Our null hypothesis is that in the population there is no difference between the Labour share of the vote and the Conservative share of the vote.

Confidence intervals for the difference between two proportions

Although it can be interesting to make estimates about the true level—or proportion—of a particular variable in the population, it is often more interesting to know about change or comparison. For example, we might be interested to see if Labour support changed when Jeremy Corbyn took over from Ed Miliband or if Conservative support changed when Boris Johnson took over from Theresa May. Alternatively, we might be interested to see if trust in politicians changed after the 2009 expenses scandal. Or if attitudes towards immigration changed after Brexit.

Obviously, when we make these sorts of comparisons, we have to rely on survey data that has been asked in exactly the same way, since even small changes in question wording can influence the response (see Chapter 11). However, there are now a number of surveys that have been carried out in the same way over time. One such survey series is the Audit of Political Engagement, carried out by the Hansard Society, which tracks British public attitudes towards politics. In recent years there have been growing concerns about a rise in anti-politics sentiment. According to this view, negative news stories and a succession of scandals have undermined the reputations of MPs, who are increasingly seen as out of

Table 15.5 Are you satisfied or dissatisfied with the way MPs in general are doing their job?

	2004	2007	2010	2013	2016
Very satisfied	1	2	1	1	2
Fairly satisfied	31	29	28	22	27
Neither satisfied nor dissatisfied	26	27	24	37	31
Fairly dissatisfied	26	26	30	23	24
Very dissatisfied	10	10	14	15	12
Don't know	5	6	3	3	3
Dissatisfied (fairly + very)	36	36	44	38	36
N	1976	1282	1156	1128	1231

Source: Hansard Society, Audit of Political Engagement 7.

touch with the concerns of ordinary people. One particularly damaging such episode was the expenses scandal, which erupted in May 2009 when the *Daily Telegraph* began exposing how many MPs from all parties had been making dubious claims for official allowances and expenses (see Heath 2010). Following an explosion of outrage, almost every MP became an object of vilification, and the media soon characterized the men and woman who sat at Westminster as 'the rotten parliament'. However, others have argued that MPs have never been held in particularly high regard, and so scandals, such as the 2009 expenses scandal, just reinforced what the public thought they already knew, rather than changing their opinion. To help us look at these different hypotheses, we can compare public attitudes towards MPs between 2004 and 2016 (see Table 15.5).

From the data we can see that the percentage of people who were very satisfied with the way in which MPs do their job was very low, at just 1 per cent or 2 per cent. By contrast the percentage of people who were very dissatisfied with the way in which MPs do their job was noticeably higher, and varied from between 10 per cent and 15 per cent. This supports the idea that MPs are generally not seen in a very positive light. If we look at the combined percentage of those who said that they were dissatisfied (either fairly or very) with the ways in which MPs in general are doing their job, we can see that the percentage increased from 36 per cent in 2004 and 2007 to 44 per cent in 2010, just after the expenses scandal, and then decreased somewhat to 38 per cent in 2013 and 36 per cent in 2016. There is not then very much evidence to support the idea that people have become more dissatisfied with MPs over time, as the level of dissatisfaction at the end of the series is almost exactly the same as it was at the beginning of the series. Nor is there much evidence to suggest that the expenses scandal did any serious long-term damage to public opinion of MPs. However, there is some evidence that levels of dissatisfaction might have been somewhat higher in the immediate aftermath of the expenses scandal in 2010 than it was at other times.

Between 2007 and 2010 the level of dissatisfaction with MPs increased by about 8 percentage points. From this it looks like there may have been a slight change in the wake of the expenses scandal. But is this change real or could it just be the result of sampling variation? Maybe the true level of dissatisfaction in the population has always been around 40 per cent, and some of the samples have underestimated it a little bit,

and the sample in 2010 overestimated it a little bit. This is a distinct possibility, and one that we would want to rule out before inferring that any real change had taken place in public perceptions about how well (or how badly) British MPs are doing their job. To try and distinguish between sampling variation and 'real' variation we use tests of statistical significance. This draws on the idea of confidence intervals discussed earlier. We can estimate a confidence interval for dissatisfaction with MPs in 2007 and 2010. The working is shown in Box 15.6.

Note that we can express the confidence intervals in terms of percentages or proportions. This time we express them in terms of percentages. The confidence interval for the level of dissatisfaction with MPs in the 2007 survey is {33.4, 38.6} and the confidence interval for the level of dissatisfaction in the 2010 survey is {41.1, 46.9}. We can see that the two confidence intervals do not overlap. The top band of the interval in 2007 is lower than the lowest band of the interval in 2010. This suggests that there has been some real change. However, we can go one step further than this, and make an estimate about how much change we think has taken place. Between the two surveys we observed an increase in levels of dissatisfaction of eight percentage points. What we want to know is that if there had been no real increase in the population, how likely is it that we would have drawn two random samples in which we observed an increase of eight percentage points?

The way we do this is to create a confidence interval for the difference between our two proportions. Our sampling distribution is now the difference between the two sample proportions. This difference, as per the central limit theorem, will be normally distributed. We know that 95 per cent of the time the 'true' difference in the population will lie within 1.96 standard errors of the sample difference, and what we want to know is the likelihood of this range including 0, meaning that there has been no change. We can calculate this in much the same way as we did before, but the only slight difference is in how we calculate the standard error of the sampling distribution. To do this we take into account both samples (S1 and S2). The formula for the standard error of the difference between two proportions is shown in Box 15.7.

BOX 15.6 Public Dissatisfaction with MPs in 2007 and 2010, Confidence Intervals

Confidence interval for survey in 2007:

$$= 36 \pm 1.96\sqrt{\frac{36*(100-36)}{1282}}$$
$$= 36 \pm 2.63$$
$$= \{33.4, 38.6\}.$$

Confidence interval for survey in 2010:

$$= 44 \pm 1.96\sqrt{\frac{44*(100-44)}{1156}}$$
$$= 44 \pm 2.86$$
$$= \{41.1, 46.9\}.$$

BOX 15.7 How to Calculate the Standard Error of the Difference between Two Proportions

The formula for the standard error of the difference between two proportions is:

$$SE_d = \sqrt{\frac{P_{s1}(1-P_{s1})}{n_1} + \frac{P_{s2}(1-P_{s2})}{n_2}}$$

$$SE_d = \sqrt{\frac{(36 * 64)(44 * 56)}{1282 + 1156}}$$

$$SE_d = \sqrt{\frac{4768}{2438}}$$

$$SE_d = \sqrt{1.96}$$

$$= 1.40.$$

Having obtained an estimate of the standard error of the distribution, we can now calculate the confidence interval in the same way:

$$ci = 8 \pm 1.96 * 1.40$$
$$= 8 \pm 2.74$$
$$= \{5.3, 10.7\}$$
$$= 8 \pm 2.74$$

We can see our confidence interval ranges from 5.3 to 10.7. This is how much change we think there has been in the level of dissatisfaction with MPs between 2007 and 2010. We can see that 0 is not contained within this confidence interval. Our estimate is that there has been change of somewhere between 5.3 percentage points and 10.7 percentage points. We are confident that this represents real change. And we will be correct in this estimate 95 per cent of the time. These results indicate that satisfaction with MPs was significantly lower in the immediate aftermath of the expenses scandal, but that the scandal did not appear to have much of a long-term impact.

Confidence intervals for means

With ordinal and interval data, we can estimate the margin of error associated with our sample means using the same general logic. To do this we calculate a statistic called the standard error of the mean. This is based on the standard deviation of the sample. As before, we can say that 95 per cent of the time the true value of the population will fall within plus or minus 1.96 standard errors of the sample mean (see Box 15.8).

These confidence intervals are particularly useful if we are interested in whether things have changed over time or not. Suppose we are interested in the ideology of the electorate in Europe (as we discussed previously), and whether there has been any real movement

BOX 15.8 How to Calculate the Confidence Interval for a Mean

The formula for the standard error of a mean is:

$$SE_{\bar{x}} = \sqrt{\frac{s}{N}}$$

where

$SE_{\bar{x}}$ = the standard error of the mean;
s = the sample standard deviation;
N = the sample size.

The 95 per cent confidence interval can then be calculated as:

$$ci = \bar{X} \pm 13.96\sqrt{\frac{s}{N}}.$$

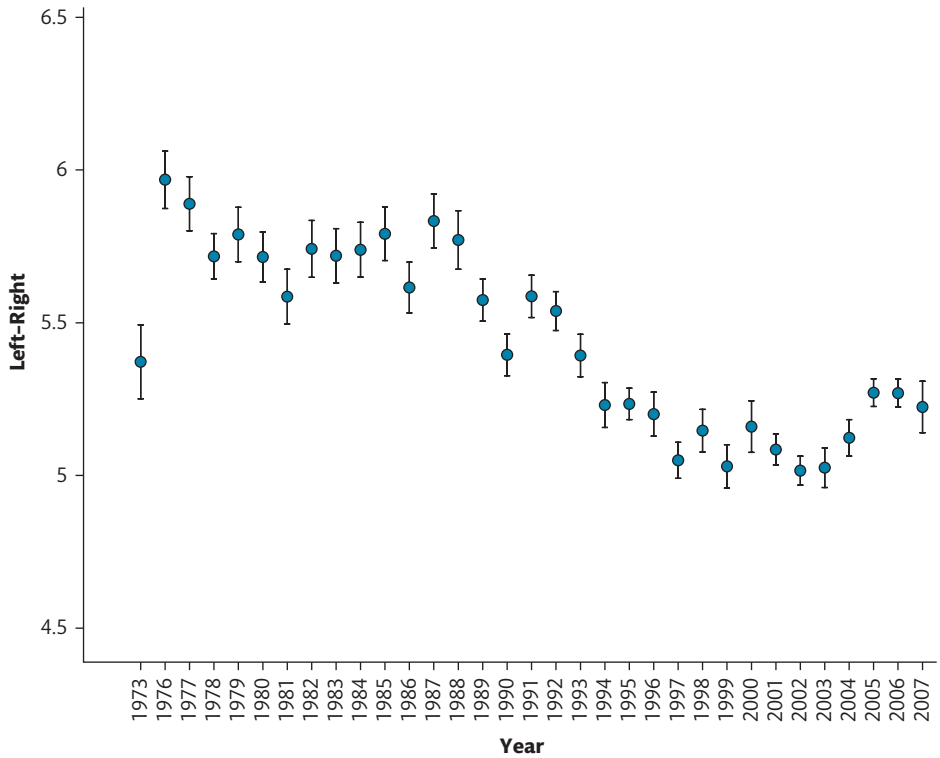


Figure 15.8 Left–right self-placement in Europe—mean position of adults with 95 per cent confidence intervals?, 1973–2016

over time. Using Eurobarometer, which has asked the same question since 1973, we can calculate the confidence interval for each sample estimate and compare them to each other. This is shown in Figure 15.8. The circles indicate the mean and the 'whiskers' indicate the 95 per cent confidence interval for the mean.

The first thing to notice is that the first point in the time series appears to be something of an outlier. It may be that between 1973 and 1976 Europeans underwent a dramatic shift to the right, but it is probably more plausible to put the anomalous estimate of 1973 down to some sort of sampling error. Since then, we can see that there have been some fluctuations up and down. By and large, the confidence intervals for two adjacent estimates tend to overlap, indicating little short-term change, but over the long run we can see that there have been some significant changes. For example, Europeans were significantly more to the left in the years immediately after the financial crisis in 2008 than they were in the years immediately before. The advantage of utilizing confidence intervals is that we are less likely to over-interpret small changes, and so have a more robust basis for assessing differences. Analysis of these differences over time can be more fully explored using time series analysis.

Conclusions

This chapter has provided an introduction to the first steps in analysing quantitative data. In doing so, we have encountered some of the building blocks of quantitative analysis: the mean, the standard deviation, and the central limit theorem. The latter provides the link between samples and populations which allows us to make inferences. Description is an important part of political research. It enables us to set the scene and describe the nature of the problem that we may then seek to go on to explain. Descriptive inference allows us to link the sample to the population. It not only enables us to make generalizations, but also allows us to assess whether things have changed or whether there are significant differences between two sample estimates. This is valuable when we move towards examining how things change over time or vary between countries. However, this univariate descriptive analysis is also often just the first step in the analysis. Often we are interested in developing explanations about why things vary or why they differ and this takes us towards developing more complicated models, which include different explanatory factors. We turn to this in Chapter 16.

Questions

1. What is the difference between description and explanation?
2. How can quantitative methods be used to describe different political phenomena? What examples can you think of?
3. What makes for good description?
4. What is the difference between description and inference?
5. What are the different levels of measurement? Give examples.
6. What is missing data? What problems can it cause? What can be done about it?
7. How do we decide what is typical? What are the different measures of central tendency and how do they differ from each other?
8. When is it appropriate to use the median? The mode? And the mean?
9. What do we mean by a confidence interval?
10. What issues influence our ability to make inferences?

Guide to Further Reading

Agresti, A. and B. Finlay (2009), *Statistical Methods for the Social Sciences*, 4th edition (Upper Saddle River, NJ: Pearson-Prentice Hall).
Covers the basics of statistical description and inference, as well as more advanced topics on regression methods, including multiple regression, analysis of co-variance, logistic regression, and generalized linear models.

Carlson, J. and M. Hyde (2002), *Doing Empirical Political Research* (Boston, MA: Houghton Mifflin).
Provides an introduction to quantitative analysis and covers some more advanced topics, with lots of examples from political research.

de Vaus, D. (2002), *Surveys in Social Research* (London: Routledge).
A detailed and accessible introduction to quantitative methods, with easy-to-read chapters on how to carry out analysis and interpret the results.

Field, A. (2017), *Discovering Statistics Using SPSS*, 5th edition (London: Sage Publications).
Provides an easy-to-follow introduction to quantitative analysis and step-by-step instructions on how to use the SPSS statistical software package to do your own analysis.

Johnson, J. B., H. Reynolds, and J. Mycoff (2007), *Political Science Research Methods*, 6th edition (Washington, DC: CQ Press).
Provides an introduction to quantitative analysis and covers some more advanced topics, with lots of examples from political research.

Rowntree, D. (2000), *Statistics without Tears: An Introduction for Non-Mathematicians* (Harmondsworth: Penguin).
An easy-to-read introduction to statistical methods.

References

Agresti, A. and B. Finlay (2009), *Statistical Methods for the Social Sciences*, 4th edition (Upper Saddle River, NJ: Pearson-Prentice Hall).

Bartle, J., S. Dellepiane, and J. A. Stimson (2011), 'The Moving Centre: Preferences for Government Activity in Britain: 1950–2005', *British Journal of Political Science* 41: 259–85.

Bischof, D. and M. Wagner (2017), 'What Makes Parties Adapt to Voter Preferences? The Role of Party Organisation, Goals and Ideology', *British Journal of Political Science* 1–11.

Bischof, D. and M. Wagner (2019), 'Do Voters Polarize when Radical Parties Enter Parliament?', *American Journal of Political Science*.

Clarke, H., D. Sanders, M. Stewart, and P. Whiteley (2004), *Political Choice in Britain* (Oxford: Oxford University Press).

Clarke, H., D. Sanders, M. Stewart, and P. Whiteley (2009), *Performance Politics and the British Voter* (Oxford: Oxford University Press).

Fieldhouse, E., J. Green., G. Evans., H. Schmitt, C. van der Eijk, J. Mellon, and C. Prosser (2017), British Election Study 2017 Constituency Results file, version 1.0.

Fieldhouse, E., J. Green., G. Evans., H. Schmitt, C. van der Eijk, J. Mellon, and C. Prosser (2018), 'British Election Study, 2017: Face-to-Face Post-Election Survey' [data collection]: http://doi.org/10.5255/UKDA-SN-8418-1

Franklin, M. (2004), *Voter Turnout and the Dynamics of Electoral Competition in Established Democracies since 1945* (Cambridge: Cambridge University Press).

Gleditsch, K. S. (2007), 'Transnational Dimensions of Civil War', *Journal of Peace Research* 44(3): 293–309.

Gleditsch, K. S. and M. D. Ward (2006), 'Diffusion and the International Context of Democratization', *International Organization* 60(4): 911–33.

Heath, O. (2010), 'The Great Divide: Voters, Parties, MPs and Expenses', in N. Allen and J. Bartle (eds), *Britain at the Polls 2010* (London: Sage), 120–46.

King, G., R. Keohane, and S. Verba (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ: Princeton University Press).

Przeworski, A., M. Alvarez, J. Cheibub, and F. Limongi (2000), *Democracy and Development: Political*

Institutions and Well-Being in the World, 1950-1990 (Cambridge: Cambridge University Press).
Wilkinson, S. (2004), *Votes and Violence: Electoral Competition and Ethnic Riots in India* (Cambridge: Cambridge University Press).
Wlezien, C. (1995), 'The Public as Thermostat: Dynamics of Preferences for Spending', *American Journal of Political Science* 39(4): 981–1000.

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

16 Patterns of Association: Bivariate Analysis

Chapter Summary

This chapter examines the association between two variables, and builds upon Chapter 15 by introducing ideas about association and causality. In doing so, it provides a link between description and explanation. In particular, we focus on hypothesis testing and significance tests; and how we can describe the pattern of association between two variables. We introduce two of the most widely used statistical analysis techniques in political research, ordinary least squares (OLS) regression and cross-tabulation. The chapter considers:

- bivariate analysis;
- cross-tabulation;
- significance testing;
- null hypothesis;
- chi square;
- correlation.

Introduction

In Chapter 15 we examined the different ways in which we can describe the distribution of one variable using **univariate statistics**. This chapter builds on these ideas. It is one thing to show that people vary in terms of their left–right ideology or environmental awareness, or that some countries are more democratic or more prone to war than others; it is quite another thing to try and explain or account for these variations. Why are some people more concerned about the environment than others? Is it because of their age? Their class? Their education? Why are some people more left wing than others? If we find that working-class people are more left wing than middle-class people, we might say that there is an association between class and left–right attitudes. If we are interested in explaining the causes of ideology, then we might hypothesize that class is an important factor. But we might also be interested in exploring why this is the case, and unpicking the causal mechanisms. Why are working-class people more left wing? Is it because of their location in the labour market? Is it because of their socialization experiences? Or is it simply a function of political mobilization and persuasion by the parties that they support?

Bivariate analysis allows us to explore these issues, and examine how two variables are related. Bivariate analysis includes a number of different methods. The type of method that is appropriate depends upon the level of measurement of our variables. As we discussed in Chapter 15, we use different methods for examining the relationship between categorical variables (nominal or ordinal) and continuous variables (interval). When both our variables are interval level, we can use ordinary least squares (OLS) **regression** (or correlation); and when both our variables are categorical, we can use **cross-tabulations** (cross-tabs). When our dependent variable is interval and our independent variable is categorical, we can do a comparison of means, as discussed in Chapter 15. Although the methods used are somewhat different, they serve the same basic purpose. They allow us to answer the following questions: Is there a relationship between two variables? And if so, what form does the relationship take? How can we describe it?

The chapter is divided into three main sections. The first section provides an introduction to the logic of bivariate analysis and discusses the difference between association and causality. The next section examines the association between two interval-level variables using correlation and regression; and the final section examines the relationship between two categorical variables using cross-tabs.

The principles of bivariate analysis

Bivariate analysis is an essential tool for getting to know your data, and helping you to identify patterns of association between two variables. The variable that refers to the phenomenon that we are trying to explain is often known as the dependent variable (or **response variable**). The variable that we think is important for explaining the dependent variable is known as the independent variable, or **explanatory variable** (see Chapter 5). So, for example, if we are interested in ethnic conflict in India, our dependent variable may be whether a particular region experienced a violent clash between Hindus and Muslims (or not) within a certain period of time (say five years). This is the thing that we want to explain. Why are some regions more prone to conflict than others? Our independent variable refers to a factor that we think may be important for explaining this. According to Ashutosh Varshney's (2002) work on the subject, one important factor refers to the structure of civil society and, he argues, regions where there are dense intra-ethnic social networks that link different communities together are less prone to conflict than regions where the two communities are more socially segregated. We can specify this hypothesized relationship as a diagram:

$$X \rightarrow Y$$

where:

Y is level of ethnic conflict in a region (our dependent variable); and X is level of intra-ethnic civil society in a region (our independent or explanatory variable).

Bivariate analysis is a way of establishing whether or not there is indeed a relationship between two variables, as the theory predicts. It allows us to compare our theoretical expectations against evidence from the real world to see if the theory is supported by what

we observe or not. In its simplest bivariate form, we have one dependent variable (which we are interested in trying to explain), such as ethnic conflict, and one independent variable (which we think influences, or causes, or helps to explain in some way our dependent variable), such as the structure of intra-ethnic civil society. When we have more than one independent variable (i.e. more than one explanatory factor), we move from bivariate analysis to multivariate analysis (which we deal with in Chapter 17), but many of the analytic principles are much the same.

The hard work in quantitative analysis is not to be found in carrying out the analysis itself. This is relatively straightforward, and once a few general techniques have been acquired, they can be applied in a wide variety of ways. Rather the hard work is theoretical, developing hypotheses and interesting research questions that engage with prior research, but also take it in new directions (see Chapter 4).

Regardless of which technique we use (cross-tabs or regression), the principles of bivariate analysis are much the same. Indeed, the principles of quantitative methods in general are much the same. We start off with a research question—or idea. That is, we seek to describe, or explain, or predict some social or political phenomenon (our dependent variables), and we have certain theoretical expectations about what factor (or factors) may be important in helping us do this (our independent variables) (see Chapters 4 and 5). We then collect data and develop indicators to measure our different concepts (see Chapters 6, 7, and 11). Then, finally, we analyse the data in order to establish whether our theoretical expectations are supported by what we observe in the real world or not. This last part is relatively straightforward, and involves obtaining answers to three basic questions.

Is there a relationship between the variables?

The first question that we want to answer is whether the variables are related to, or associated with, each other or not. Two variables are associated if knowing something about one variable helps us to know something about the other variable. In formal terms, we can say that variables are associated if the conditional distribution of one variable varies across the levels of the other. For example, if economically underdeveloped countries tend to be authoritarian and economically developed countries tend to be democratic, we could say that economic development and democracy are associated; that is, the level of democracy varies across levels of economic development. In more formal terms, we can assess whether the sample associations that we observe are statistically significant or not.

Underpinning significance testing is the idea of hypothesis testing, and the **null hypothesis**. The null hypothesis is the hypothesis that we empirically test, and is basically the hypothesis of no association (hence the null). In this sense, it has parallels with Karl Popper's 'falsificationist thesis', which we discussed in Chapters 2 and 3. This principle of falsification lies at the heart of hypothesis testing in quantitative methods. We do not therefore test to see if there is a relationship between two variables, but rather we test to see if we can reject (or falsify) the hypothesis with which there is no relationship. So, for example, if we are interested in the relationship between economic development and

democracy, we may think that rich countries tend to be more democratic than poorer countries. We do not directly test this though. Rather, our null hypothesis is that there is no association between economic development and democracy. If we find evidence to reject the null hypothesis, then we can reject the idea that there is no relationship and infer that there probably is a relationship. Whether we reject (or fail to reject) the null hypothesis is incredibly easy to decide. We will discuss this in more detail later in the chapter, but the standard practice is simply to reject the null hypothesis if something called the p value is less than 0.05. The p value refers to the probability that the null hypothesis is true. So a p value of 0.05 tells us that there is a 5 per cent chance that the null hypothesis is true. If the probability of it being true is very low (by convention less than 5 per cent), then we assume that it is probably false.

Null findings

Sometimes people think that the goal of research is to uncover new relationships, and so they get very disappointed if they can't reject their null hypothesis. However, this concern is misplaced. The goal of research is to answer interesting questions. If the question is important and the research design is sound, then null findings can be just as important and interesting as statistically significant findings. Indeed, null findings can be incredibly important in their own right. For example, in health research the finding that there is no link between vaccination jabs and autism is incredibly important, as many people were under the mistaken impression that there was such a link and so stopped vaccinating their children, increasing the spread of infectious diseases. This null finding therefore helped to save many lives. There are also many instances of important null findings in political science. Indeed political scientists often theorize that particular variables should not matter. For example Svolik (2008) argues that economic development does not affect the timing of authoritarian reversals and Fearon and Laitin (2003) argue that ethnic and religious diversity do not make civil wars more likely. And sometimes we simply want to know the answer to an interesting question. In these circumstances the important thing is that we do our best to answer the question with the most appropriate research design and best data available. If the results at the end of that are null, we still learn something important about that world.

How can we describe the relationship between the variables?

If we find evidence to suggest that two variables are statistically associated, we could then go on to describe the pattern of this association. Principally, we are interested in whether the pattern of the relationship conforms to our theoretical expectations or not (i.e. that richer countries are more democratic). How we describe this association varies according to the method that we use (cross-tabs or regression). With cross-tabs we describe the association with reference to percentages, and with regression we describe the association with reference to something called the **line of best fit**. But in either case we are principally concerned with whether the pattern that we observe is in the expected direction predicted by our theory.

How good is our explanation?

Finally the third question that we need to bear in mind is to do with how good—or how robust—we think our findings are. Put simply, are they convincing? On what grounds could they be challenged? Invariably, with bivariate analysis our explanation will be partial since we are considering the impact of only one variable on the dependent variable. We then need to think about what other variables or factors may be important, and whether our results still stand up when we take these into account. This we deal with in Chapter 17.

Association and causality

In many cases, identifying patterns of association between two variables helps us to make claims about causality. We can think about this in deterministic terms or in probabilistic terms (see Chapter 5). However, it must be remembered that just because two variables are associated with each other, it does not mean that the relationship is causal. And trying to demonstrate that a relationship is causal is not straightforward. A causal relationship can be thought of as one where *X*—our independent variable—causes *Y*—our dependent variable. For example, does economic development influence (or cause) democracy? Does ethnic diversity influence (or cause) civil war? For a relationship to be considered causal, we need to establish three conditions. First, there must be statistical association. If two variables are not associated with each other, then there cannot be a causal relationship. However, association by itself is not enough to suggest causality. It is a necessary but not sufficient condition. Second, there must be the correct temporal ordering. For variable *X* to cause variable *Y*, *X* must be temporally prior to *Y*. Sometimes this is easy to establish. It is clear, for example, that a person's age is temporally prior to their decision to go out to vote. We could not argue that abstaining makes you younger. But at other times, it is less clear-cut, and it is more difficult to disentangle the temporal ordering or direction of causality. For example, does economic development make a country more democratic? Or could it be the other way round? Does being democratic make a country richer? The theory may be uncertain, and questions of this type bedevil empirical research. Does *a* cause *b* or does *b* cause *a*? Or indeed is there a reciprocal relationship?

Third, we must rule out alternative explanations. We will return to this point in Chapter 17 when we consider multivariate analysis. However, even when these requirements are all met, we can still only make tentative claims about causality, and perhaps what we have achieved can better be described as 'robust association'. Ultimately, then, even if we develop theoretical hypotheses that are causal in nature, when we come to analysing the data, all we can do is see whether the pattern of association is consistent with our expectations or not.

Data considerations for bivariate analysis

Bivariate analysis includes a number of different methods. The type of method that is appropriate depends upon the level of measurement of our variables. In the following sections, we examine the association between two interval-level variables using OLS regression, and then consider the association between two categorical-level variables using cross-tabs.

Bivariate analysis for interval data

There are a number of ways in which we can examine the relationship between two interval-level variables. The most common technique is simple linear regression (OLS regression), and Pearson's correlation. Both techniques are closely related, and we discuss them in detail later in the chapter. The workhorse for both techniques is the scatter plot. Bivariate analysis with interval-level data can best be depicted through the use of figures (at least when there is a small number of cases), as this creates a clear visual representation of how the two variables are related (or not). The pattern of this relationship can then be summarized using statistics. This section focuses on simple linear regression, which is one of the most widely used statistical techniques in political research, and which can be used both for bivariate and multivariate analysis. The first part of this section examines how scatter plots can be used to depict the relationship between two interval-level variables. The second part then looks at how we can summarize this relationship using the line of best fit—otherwise known as the **regression line**. The third part then moves on to questions of significance, and discusses how we can make inferences about whether the sample association exists in the wider population.

Bivariate analysis using scatter plots: describing the relationship

Figures provide an intuitive way of examining the relationship between two interval-level variables when we have relatively few cases. They visually represent the important characteristics of the relationship, and just by eye-balling the scatter plot, we can immediately get a rough idea of whether there is a relationship or not, and if there is, whether it is positive or negative and whether it is weak or strong. In this section, we examine how figures can be used to depict the association between two interval-level variables, and how this association can be summarized using the line of best fit. The line of best fit is also known as the regression line, which forms the basis of simple linear regression analysis, and which we will examine in more detail.

To illustrate these techniques let us consider the relationship between the level of economic development of a country and its level of democracy. A great deal of research has examined this topic, and the research has gone in many different and sophisticated directions, but the starting point for much of the analysis is the simple observation that rich countries tend to be more democratic than poor countries. To examine this relationship we need to develop empirical indicators for each of our variables. To measure economic development is fairly straightforward. There are a number of measures we could use, but perhaps the most straightforward is simply gross domestic product (GDP) per capita. To measure a country's level of democracy is slightly more problematic (see Chapter 7). Democracy is a difficult concept to define, and wherever there is definitional ambiguity, there is also likely to be controversy over measurement. There are a number of widely used indicators to measure democracy, from simple dichotomies of whether a country is democratic or not, based on whether it holds free and fair elections (Przeworski et al. 2000), to more extensive scales based on political rights (Polity IV; Marshall and Jaggers 2010) and civil rights (Freedom House 2010). For this example we opt for Polity IV, since it distinguishes between different levels of democracy (from very authoritarian to fully democratic) on a scale of 0 to 10 (where 10 is the most democratic).

The next step is to collect information on different countries for each of these variables. Each country therefore represents a case, and for each case we collect information on its level of democracy (measured by its Polity IV score) and its level of economic development (measured by GDP per capita in US dollars). To simplify this example we start by considering just ten countries. The data for these ten countries are shown in a data matrix in Table 16.1, similar to how data are stored in most statistical software packages. Each row represents a case, and each column represents a variable.

From the raw data, it is difficult to ascertain whether there is much of an association between a country's score on each of the variables. But this becomes much clearer if we represent the data graphically. Our expectation is that those countries with low levels of economic development will also have low levels of democracy, and that those countries with high levels of economic development will have high levels of democracy. To see if this expectation is fulfilled we can simply plot the values for each variable against each other on a scatter plot. These values can be thought of as the coordinates for each country, which form the basis of correlation and regression. This is done in Figure 16.1, where the horizontal axis (the x-axis) refers to a country's level of economic development and the vertical axis (the y-axis) refers to a country's level of democracy. Each dot represents a country, and its position on the scatter plot relates to its score on each of the variables. Just from eye-balling the results, we can see that our expectation appears to be confirmed. The countries with relatively high levels of economic development, like Costa Rica and Mexico, also tend to have relatively high levels of democracy, and the countries with relatively low levels of economic development, like Chad and Cambodia, tend to have relatively low levels of democracy.

We can phrase this in a slightly different way and say that there appears to be a positive relationship between the level of economic development of a country and its level of democracy. That is, countries with *higher* levels of economic development tend to have *higher* levels of democracy. We can see this quite clearly from the scatter plot, but we can also provide a summary measure of this relationship by fitting to the data something commonly known as the line of best fit, or the regression line (see Figure 16.2).

Table 16.1 Democracy and development, data matrix

Country	*X*–GDP per capita	*Y*–Level of democracy
Chad	$871	1
Cambodia	$1,446	3
Georgia	$2,664	5
Ecuador	$3,203	6
Paraguay	$4,426	7
Bulgaria	$5,710	8
Venezuela	$5,794	7
Brazil	$7,625	8
Costa Rica	$8,650	10
Mexico	$9,023	8

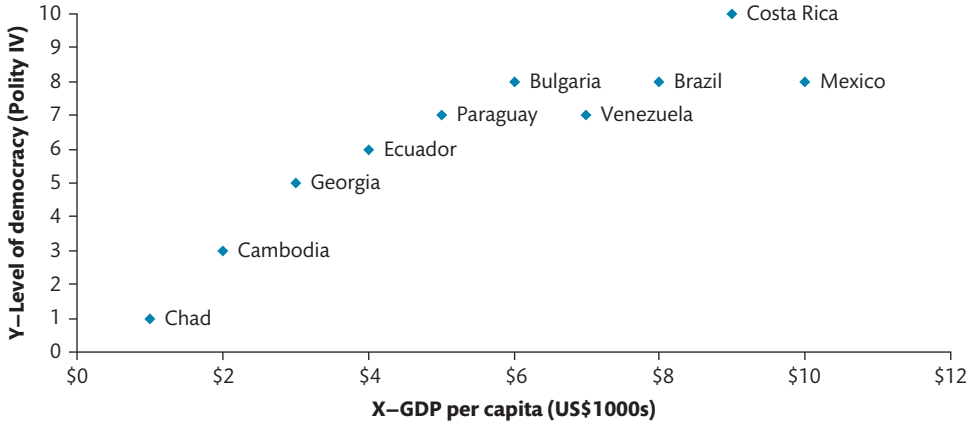


Figure 16.1 Scatter plot of economic development and level of democracy

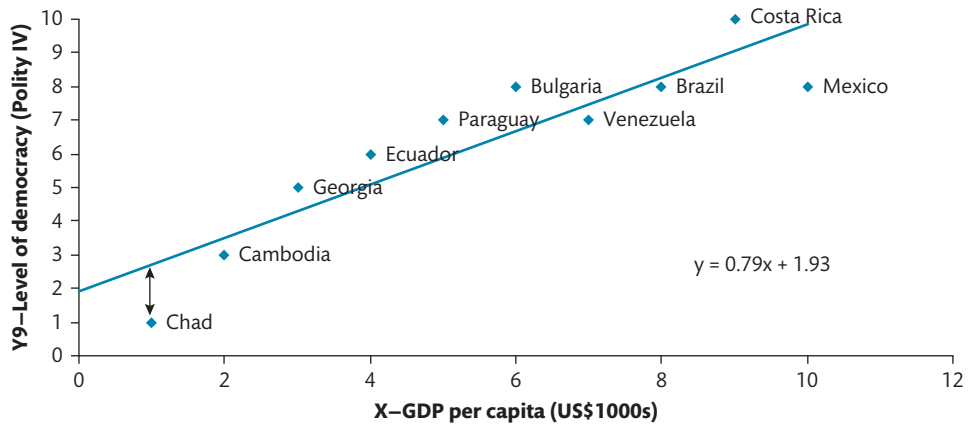


Figure 16.2 Scatter plot of economic development and level of democracy, with line of best fit

The regression line gives a neutral, impartial way of identifying the best fitting line to a scatter of dots in a two-dimensional space. This method involves identifying a line that minimizes the sum of squared vertical deviations from itself. The deviations are the distance between the line of best fit and each observation. The regression line cuts through the middle of all the observations. Some of the observations are above the line, like Bulgaria; some of the observations are below the line, like Chad; but the best fitting line minimizes the distance shown by the arrow between the line and the observations (also known as the residuals). This procedure is called least squares and is often called OLS. The calculation of the line of best fit is not particularly interesting from a substantive point of view, and there is no need to go into it in detail, since practically all software packages will fit the line for you. But the general principle is simple enough. The important thing is to be able to interpret the line once it has been fitted.

The regression line can be represented with the following equation:

$$Y = a + bX$$

where:

Y is the dependent variable (the one on the vertical axis);
X is the independent variable (the one on the horizontal axis);
a is the intercept;
b is the slope.

Sometimes this equation is written with a 'hat' over the Y to indicate that the dependent variable is not an exact function of X, and is a predicted value for which there is associated error (meaning that not all of the observations fit perfectly onto the line of best fit). Either way, the main points of interest are the values for the intercept (sometimes called the constant) and the slope coefficient and how these can be interpreted.

From the example earlier, the values that we obtain for our line of best fit can be expressed as follows:

$$Y = 1.93 + 0.79X.$$

The **intercept**, or the **constant** as it sometimes known, is the point where the line of best fit crosses the y-axis. It thus refers to the value of Y when X is zero. Here, the intercept of 1.93 refers to the predicted level of democracy for a country with a GDP per capita of zero. In most cases, this value is not of much analytical interest, since a value of zero on the independent variable does not make much theoretical or empirical sense. We are therefore rarely interested in discussing the size of the constant, but it is important if we are interested in prediction, and making a guess about Y for some given value of X, which we will return to later.

Of far more substantive interest is the *slope* of the line, which describes the functional relationship between x and y, and is sometimes known as the **slope coefficient** or **regression coefficient**, depicted by the letter b. The slope coefficient can tell us a number of things. First, if the slope coefficient is zero, then the slope of the line is flat, parallel to the x-axis, and there is no relationship between the two variables. In other words, an increase in the value of our independent variable has no impact on the value of the dependent variable. Second, the slope coefficient can also tell us whether the relationship is positive or negative: a positive value for the slope coefficient indicates that as values of X increase, values of Y also tend to increase, whereas a negative value for the slope coefficient indicates that as values of X increase, values of Y tend to decrease. And, third, it also allows us to quantify the magnitude of this increase (or decrease) and specifically it tells us the number of units that Y changes for each unit change in X. So, in this example, a slope coefficient of 0.79 tells us that for every unit increase in X of $1,000 (since GDP per capita is measured in thousands of dollars), the predicted or estimated level of democracy in a country increases by 0.79 points on the Polity IV scale (which ranges from 0 to 10).

The value of the regression coefficient might appear small, but its size is partly a function of the scale on which the independent variable (and dependent variable) is measured. For example, if GDP per capita were measured in dollars rather than thousands of dollars, the size of the coefficient would be 1,000 times smaller. It is therefore often helpful to illustrate

the substantive impact of the independent variable by calculating the predicted value of *Y* for different values of *X*. So we might want to compare the predicted level of democracy for a relatively well-off country, say one with a GDP per capita of $8,000, and for a relatively poor country, say one with a GDP per capita of $2,000. We can then plug the relevant numbers into our equation (remembering to take into account the value of the intercept) as follows:

$$Y = 1.93 + 0.79X.$$

The predicted level of democracy for a well-off country with GDP per capita of $8,000:

$$Y = 1.93 + 0.79^{*}8$$
$$Y = 1.93 + 6.32$$
$$Y = 8.25.$$

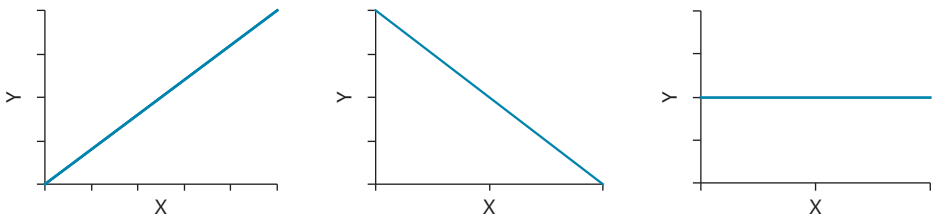
The predicted level of democracy for a poor country with GDP per capita of $2,000:

$$Y = 1.93 + 0.79^{*}2$$
$$Y = 1.93 + 1.58$$
$$Y = 3.51.$$

An increase in GDP per capita from $2,000 to $8,000 is associated with an increase in democracy of 4.8 points, from 3.5 to 8.3. Or, to put it another way, the predicted level of democracy in a relatively well-off country is more than twice as high as the predicted level of democracy in a relatively poor country.

Is there a relationship? Significance testing with interval variables

For two variables to be associated, then, the slope of the line must be different from zero. If the slope is positive, then we observe an upward-sloping line, as shown in Figure 16.3(a), indicating that as the value of the independent variable increases, the predicted value of the dependent variable also tends to increase. If the slope is negative, then we observe a downward-sloping line as shown in Figure 16.3(b), indicating that as the independent variable increases, the predicted value of the dependent variable decreases. If the slope coefficient is zero, however, then the line of best fit is flat, as shown in Figure 16.3(c), indicating that as the independent variable increases, there is no linear change in the predicted value of the dependent variable.



(a) Positive relationship (b) Negative relationship (c) No relationship

Figure 16.3 Different types of linear relationship

This raises the question of what counts as a slope coefficient that is different from zero. How big does the coefficient need to be? If our slope coefficient is, say, 0.2, can we infer that there is a real relationship? To have confidence that the relationship really exists, we have to establish how likely it is that the slope also exists in the population. To help us do this we draw on the principle of significance testing, which builds directly on what we discussed in Chapter 15.

So far we have used simple linear regression as a descriptive tool to describe the pattern of the relationship between two variables in our sample. But linear regression modelling is more powerful as a tool of inference: it allows us to make estimates of the parameters for the model *in the population*. We use inferential statistics because we want to say something about the population from which our sample is drawn. That is, in the world at large, do rich countries tend to be more democratic than poorer countries?

In order to answer this question we carry out a significance test. The method we use to test for significance in regression is called the **t test**. The t test (or t ratio) enables us to assess how likely it is that the sample association we observe could just be the result of **sampling variation**. Even if we had data on all the countries in the world, we would still treat this as a sample, since what we are really interested in is the relationship between the variables, and the countries in the world are in a sense a sample of the values these variables can take. The only parameter of the simple linear regression model for which we will describe methods of statistical inference is the slope coefficient *b*. Tests and confidence intervals for population values of the intercept *a* are almost never substantively interesting, so they will not be considered here. Significance testing involves a number of steps. The first step is to clarify exactly what it is that is to be tested. This is known as stating the null hypothesis. The second step involves making a decision on whether the null hypothesis can be rejected or not. To do this we refer to something that is known as the test statistic (t-statistic; or t ratio), which provides a summary measure of the difference between what we observe and what we would expect to observe if there were no relationship. Finally, we make a decision on whether this difference is statistically significant or not, with reference to the p value, which refers to the probability of obtaining the observed t value if the null hypothesis were true. The standard practice is to reject the null hypothesis if the p value is less than 0.05. We discuss each of these steps later.

The null hypothesis is that there is no linear relationship between the explanatory variable *X* and the response variable *Y* in the population. This corresponds to a slope coefficient with a value of zero. Graphically, this corresponds to a regression line in the population which is flat. The null hypothesis can thus be expressed in words as:

H_0: There is no linear association between *X* and *Y* in the population.

On a point of clarity, we do not ever prove a null hypothesis to be true or false. Rather we collect evidence to make an informed decision based on the balance of probability. Because we are dealing with probabilities rather than absolutes, there is always a risk that we make an error. In particular, there are two main types of error that we can make. These are known as the **Type I error** and the **Type II error**. A Type I error is when we incorrectly reject a null hypothesis that is actually true. The way to reduce these errors is just to raise the threshold for failing to reject the null hypothesis, from the 0.05 level to the 0.01 level. The problem with this approach though is that it risks introducing what is known as the Type II error,

which is failing to reject a null hypothesis that is actually false. Type I errors are conventionally considered more serious than Type II errors, so what we most want to avoid is rejecting the null hypothesis unnecessarily. This implies that we will maintain the null hypothesis unless data provide strong enough evidence to justify rejecting it, a principle which is somewhat similar to Popper's injunction to 'keep a theory until falsified', discussed in Chapter 3, or even the 'innocent until proven guilty' principle of law.

Despite our dislike of Type I errors, we never try to avoid them completely. The only way to guarantee that the null hypothesis is never incorrectly rejected is never to reject it at all, whatever the evidence, which is obviously not a sensible decision rule for empirical research. Instead, we decide in advance how high a probability of Type I error we are willing to tolerate, and then use a test procedure with that probability. The convention is that we are prepared (if not happy) to be wrong 5 per cent of the time, and so we use a 5 per cent level of significance to make decisions from a test. The null hypothesis is then rejected if the sample yields a t-statistic for which the p value is less than 0.05. This is our critical value.

The test statistic

To make some sensible statement about whether the slope in the population is flat or not, we need to rely upon statistical and probability theory. From our sample, we have an estimate of the population slope, but it is only that: an estimate. What we want to know is whether there is enough evidence to suggest that our estimate is significantly different from zero. If there is, we can reject the null hypothesis and infer that in the population there probably is a relationship between the two variables. But if we do not have enough evidence to do this, we fail to reject the null hypothesis. The logic underpinning the t test is quite intuitive. We have an estimate of the slope coefficient in the population. We want to know if this is significantly different from zero, but our coefficient is only an estimate. The true value may be a little higher or a little lower. To see whether it is significantly higher or lower than zero we utilize something called standard error. The standard error of the regression coefficient refers to the precision of our estimate. From the standard error we can calculate a confidence interval in much the same way as we did in Chapter 15. The 95 per cent confidence interval for the slope coefficient is calculated as the coefficient plus or minus 1.96 multiplied by its standard error. So, for example, if the slope coefficient is 3, and the standard error is 2, our 95 per cent confidence interval is between 6.92 and –0.92 (3 plus or minus 3.92). This is not a very precise estimate and since 0 is contained within that confidence interval, we cannot rule out the possibility that the real population parameter is zero. We would thus fail to reject the null hypothesis. As a shorthand, when the slope coefficient is less than twice the size of its standard error, we fail to reject the null hypothesis (this is why it is important always to report the standard errors of the regression estimates).

The t-statistic is simply the slope coefficient divided by its standard error. The formula is shown in Box 16.1.

The t-statistic can be thought of as a measure of the difference between what we would expect to see if the null were true (e.g. the slope would be zero), and what we actually observe. Larger values of t therefore indicate that our sample does not conform to the expected pattern of no association. Using probability theory, we can establish how likely it is that we would observe a t of the given value if the null hypothesis were actually true. The p value

BOX 16.1 How to Calculate the Test Statistic

The formula for the t-statistic is:

$$t = \frac{\hat{\beta}}{s\hat{e}\left(\hat{\beta}\right)}$$

where:
β is the estimated coefficient; and
$s\hat{e}(\beta)$ is the estimated standard error of the coefficient.

gives us the exact probability of drawing a sample with a t-statistic of the observed value or greater from a population in which there is no association between the two variables (e.g. the slope is zero). By convention, if the probability is lower than 0.05, we reject the null hypothesis. What we are saying is that the probability of obtaining a t value of this size if the null hypothesis is true is so small that we are confident that it must therefore be false. But this is somewhat different to proving that it is false. The key word here is confidence. We can never know for sure whether the null hypothesis is true or not. We know that ninety-five times out of a hundred it will be the correct decision to reject the null hypothesis, but we also know that five times out of a hundred we will reject a null hypothesis that is actually true.

The strength of the relationship: a question of fit

So far we have examined whether (or not) we think there is a relationship between two interval-level variables, using significance tests, and how we can describe the form of that relationship, with reference to the direction of the relationship and the magnitude of the coefficient. There is, however, a third aspect of the relationship that it is often helpful to consider. This is sometimes known as the strength of the relationship, but is perhaps better described as the fit. If we were to say that there was a strong relationship between two variables, we could understand this in two different ways. First, we could think of a strong relationship as being one where our independent variable has a 'big' impact on our dependent variable, that is, one where the slope of the regression line is very steep. But we could also think of a strong relationship as being one where our line of best fit produces very little error, that is, when all the observations are clustered very tightly around the line.

The line of best fit allows us to predict values of Y for given values of X. However, we almost always make some errors in our prediction of Y. The errors are called the residuals, and are the distance between the line of best fit and the actual observations that the line passes through. The more accurate the guesses of $\hat{Y}$, the smaller the residuals, and the better we can say the line *fits the data*. This is illustrated in Figure 16.4, which shows two graphs with exactly the same values for the intercept and slope coefficient. In the first graph, we can see that the observations are quite spread out around the regression line, whereas in the second graph, all the data points are very close to the regression line. We can describe the difference between these two regression lines in terms of which one provides the better 'fit' to the data. Both Pearson's correlation and R-square are measures for summarizing this fit.

R-square (R^2) is one of the most commonly reported statistics in quantitative political research. One way to think about it is that it provides a summary measure of how much better

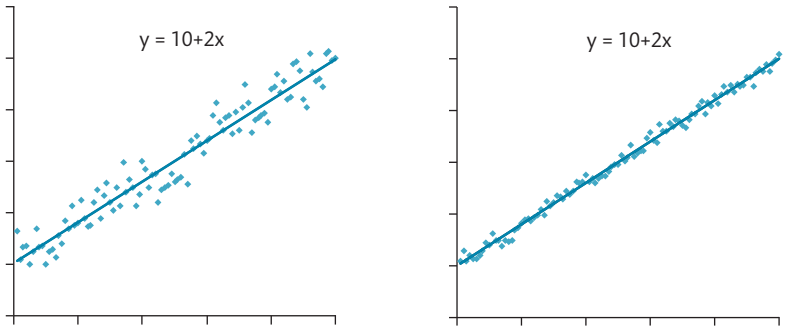


Figure 16.4 The 'fit' of the regression line

off we are in our predictions of Y using the regression line than some other estimation technique (basically, guessing). It describes the proportion of the variation in the dependent variable that our model is able to 'explain'. If we had no information about how X is related to Y, then our best guess of Y would simply be its mean value, $\overline{Y}$. As Figure 16.5 shows, for a given value of X, the value of $\overline{Y}$ may be quite a long way away from the observed value of Y_i. We can summarize this error from guessing by adding up the squared distance between each value of Y_i and the mean value of Y $(Y - \overline{Y})$. This is known as the total sum of squares (TSS), and can be regarded as a measure of the total variation in our dependent variable.

Out of this total variation, we want to know how much our model explains. Using the regression line, we make predictions of Y, denoted $\hat{Y}$. Sometimes our predictions will be good and $\hat{Y}$ will be close to the observed values of Y; sometimes our predictions will be bad and $\hat{Y}$ will be a long way away from the observed values of Y. We can break this down into what our model explains—the explained variation, and what it does not—the unexplained variation. The explained part of the model can be thought of as the difference between our predicted value of Y and the mean value of Y. This is known as the 'Model sum of squares', denoted SSM.

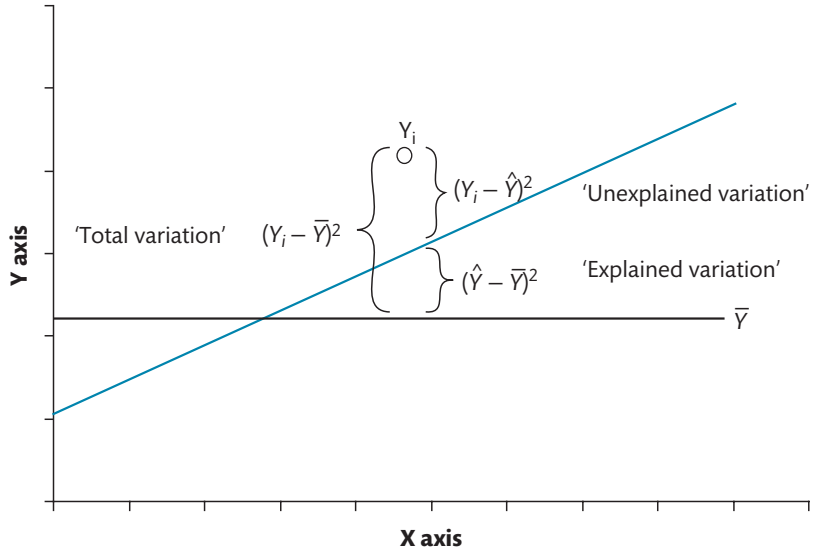


Figure 16.5 Explained and unexplained variation

The unexplained part of the model (or the error) can be thought of as the difference between our predicted value of *Y* and the observed value of *Y*. This is known as the 'Error sum of squares', denoted SSE. These sources of variation are related to each other so that the total variation of Y is the sum of the explained variation and the unexplained variation. The R-square statistic is simply defined as the proportion of the total variation that the model explains (see Box 16.2). This is known as a PRE measure (Proportional reduction in error). It is the proportion of the total variation of *Y* explained by the regression model. R-square ranges between 0, when the regression explains nothing (the slope coefficient is 0), and 1, when there is perfect association (all the observed values of *Y* fall exactly on the line of best fit). The formal interpretation of the R-square is: 'The proportion of variance in *Y* that can be explained by the variance in X'. So if we obtain an R-square of 0.50, we can say that 50 per cent of the variance in *Y* can be explained by the variance in *X*. Or, that our model explains 50 per cent of the variation in the dependent variable. This can also be put in a slightly different way, and we could say that we make 50 per cent fewer errors when guessing *Y* while knowing *X*, as compared to guessing *Y* when not knowing *X*.

Pearson's correlation

Closely related to R-square is Pearson's correlation coefficient, or Pearson's r as it is often known. When Pearson's r is squared, it becomes R-square, and can be interpreted as a PRE measure, as discussed previously. But unlike R-square, Pearson's r can also take negative values and ranges from –1 to + 1. This means it can provide a useful summary about the

BOX 16.2 How to Calculate R-Square

The formula for calculating R-square is:

R-square = SSM/TSS

where:
SSM is Model Sum of Squares (explained variation); and
TSS is Total Sum of Squares (explained variation)
These sources of variation are related to each other so that:

Total variation of Y = Variation explained + Unexplained variation by regression

TSS = SSM + SSE.

The formula for calculating these sources of variation is:

$$TSS = \sum \left(Y_i - \bar{Y}\right)^2.$$

$$SSM = \sum \left(\hat{Y} - \bar{Y}\right)^2.$$

$$SSE = \sum \left(Y_i - \hat{Y}\right)^2.$$

where:
Y_i is observed value of *Y*; and
$\hat{Y}$ is predicted value of *Y*; and.
$\bar{Y}$ is mean value of *Y*

direction of the relationship between two variables (whether it is positive or negative), as well as about the strength of the relationship (in terms of its fit to the data). When there is perfect negative association between two variables, Pearson's r is –1; when there is perfect positive association between two variables, Pearson's r is + 1; and when there is no linear association between two variables, Pearson's r is 0.

Example 1: democracy and development

To illustrate how all these steps fit together, we can return to the example of the relationship between democracy and economic development that we introduced earlier, and examine the association with real data. Does democracy vary by a country's level of economic development? If so, how? The data set that we use to answer this question is the cross-national democracy data set compiled by Pippa Norris (2015), which contains valid information on 145 countries. Each country therefore represents a case, and for each case we have information on the level of democracy (measured as Polity IV score) and level of economic development (measured as GDP per capita). We can specify the model that we estimate algebraically:

$$Y = a + b_1 X_1 + \varepsilon$$

where:

Y refers to level of democracy measured by Polity IV (our dependent variable);
X refers to GDP per capita in US$1,000s (our independent variable);
ϵ is the error term.

Our null hypothesis is that there is no association between a country's level of economic development and level of democracy. That is, the true parameter of the slope coefficient b is zero. Our first task then is to establish whether the null hypothesis can be rejected or not. To do this, we estimate the model using a statistical software package, such as SPSS. The results are presented in Table 16.2. The table contains quite a lot of information, and it is important to be clear about what each piece of data refers to. The first row in the main body of the table is labelled 'constant', and the value in the column labelled '*b* coefficient' refers to the value of the constant—or the intercept. This simply tells us the point at which the line of best fit crosses the y-axis. Although most software packages will also provide additional information, such as the standard error, the t-statistic, and the probability, none of these are of much interest with reference to the intercept.

Of more substantive interest is the row labelled 'GDP per capita'. Our first task is to establish whether the null hypothesis that there is no association between economic development and democracy can be rejected or not. To make a decision on this we inspect the

Table 16.2 Democracy and development, OLS regression, parameter estimates

	Unstandardized beta coefficient	Standard error	*t*	*p* value
Constant	1.96	0.61	3.20	0.002
GDP per capita	0.23	0.06	4.15	0.000

Notes: Dependent variable: polity democracy ten-point score, 2000. N = 145; R-square = 0.11.
Source: Pippa Norris (2015).

associated t-statistic and the p value. The t-statistic is 4.15, and the p value is <0.0005. The p value tells us the probability of observing a t-statistic of 4.15 or greater if the null hypothesis were true. In this instance, the probability of the null being true is very small indeed, and more importantly, it is much less than our critical cut-off point of 0.05. We can therefore confidently reject the null hypothesis of no association, and infer that there probably is an association between the two variables.

Having established that there is a significant relationship, we next turn our attention to describing the pattern of this relationship. We do this with reference to the slope coefficient, which is sometimes reported as the unstandardized beta coefficient. The slope coefficient of 0.23 tells us that there is a positive association between economic development and democracy, and, more precisely, that the predicted level of democracy in a country increases by 0.23 points on the Polity IV scale for every US$1,000 increase in a country's GDP per capita. This is consistent with our prior expectations, and the evidence does suggest that on average rich countries tend to be more democratic than poor countries.

Finally, we can assess how well this model fits the data by considering the R-square, which is reported in the note to the table. The R-square is 0.11, which tells us that 11 per cent of the variance in democracy scores can be explained by the variance in GDP per capita. On the face of it, this might not seem very much. In terms of explaining why some countries are more democratic than others, we have only scratched the surface, and there may be many other reasons apart from economic development for why some countries are more democratic than others. To gain a fuller understanding of the sources of this variation we may therefore want to explore the impact of other factors, which leads us to multivariate analysis, which we will deal with in Chapter 17.

Of course, we have to bear in mind that what we have established is a statistical association between the two variables, and this is very different from establishing a causal relationship between economic development and democracy, even if that is what the theory suggests. Statistical association is, of course, the first condition of establishing causality, but we might have serious questions about whether the other two conditions of correct temporal ordering and elimination of alternative explanations have been satisfactorily dealt with. We do not know that if a country were to become richer, it would also become more democratic (to investigate that, we would need to look at changes over time) or even why rich countries are more democratic than poorer ones. Bivariate analysis thus often leads to more questions than it can answer and constitutes the first step in a more involved investigation. We will return to some of these points in Chapter 17.

Bivariate analysis for categorical data: cross-tabulation

So far we have examined the bivariate association between two interval-level variables. However, many of the social and political phenomena that we are interested in exploring are not measured on an interval scale. Research on voting behaviour, political protest, democratic breakdown, and civil war often deal with dependent variables that are nominal, and focus on whether someone votes for one particular political party or another, or whether a country has experienced democratic breakdown or civil war or not. Indeed, democracy can also be thought of in nominal terms: either a country is democratic or it is not. To examine these

types of variables, then, we need to use a different technique. Cross-tabs are the appropriate method of bivariate analysis when you are interested in the association between two categorical (nominal or ordinal) variables. Cross-tabs work best when the variables have a limited number of categories. Too many categories make the tables unwieldy and difficult to analyse, and also run the risk of introducing sparse or empty cells (see Chapter 15). But even if variables do have a lot of categories, such as interval-level variables like age, it can often be helpful initially to split them into categories, such as age groups, and look at the cross-tabs. Although cross-tabs involve analysing the relationship between two variables in a slightly different way from the method of simple linear regression discussed earlier, the analytic principles are very similar. It enables the researcher to answer two main questions. Is there a relationship between the two variables? And if so, what does that relationship look like? How can we describe it?

Bivariate analysis using tables: describing the relationship

The essence of cross-tabs is to compare the frequency distributions (see Chapter 15) for two or more categories. These are sometimes known as conditional distributions. They are conditional because the distribution of the dependent variable depends upon the level of the explanatory variable. To illustrate the logic of cross-tabs we can think about the relationship between age and turnout. In recent years, there has been a great deal of academic research on the relatively low levels of electoral participation among young people, who are generally much less likely to vote than their older counterparts. However, in the 2017 UK General Election some people argued that there was a 'Youthquake', and that turnout among young people was substantially higher than it had been previously. This sparked a debate about what difference, if any, there was in turnout among different age groups between 2015 and 2017, and whether these differences could reasonably be interpreted as constituting a quake or not (see Research in Action Box 16.1).

To explore this issue we examine the relationship between age and turnout using survey data collected by the British Election Study (BES) in 2015 and 2017 (Fieldhouse et al. 2016; Prosser et al. 2018). Since age is an interval-level variable we need to recode it into different age bands, transforming it into an ordinal variable. By contrast, turnout is a nominal variable. People either voted or they didn't. When a variable is binary—that is, when it contains only two categories (voted/didn't vote)—it is often referred to as a **dummy variable**. We can then proceed to see whether (and how) patterns of turnout vary by age group.

We start by considering the relationship between age and turnout in 2015. Since we are interested in trying to explain turnout, this is our dependent variable. Our independent variable, or explanatory variable, is age. It is convention that the dependent variable goes along the rows, and the independent variable goes down the columns. Turnout is therefore the row variable and age is the column variable (see Table 16.3). The intersection of a row and a column is the cell of a table. The frequencies in the internal cells describe the joint distribution: they show how many cases have each possible combination of the row and column variables. The first step is just to aggregate the cell counts for each different combination of responses. For example, the 2015 BES (Fieldhouse et al. 2016) interviewed 2,945 people. We can see that overall 2,172 people voted and 773 people did not vote (see column labelled 'Total'). In addition, we can see that 462 of the youngest age group voted, compared with 769 of the middle age group and 941 of the oldest age group.

Table 16.3 Age and turnout in 2015 British General Election, cell counts

	18–39 years old	40–59 years old	60 years old and over	Total
Did not vote	367	258	148	773
Voted	462	769	941	2,172
Total	829	1,027	1,089	2,945

Source: British Election Study 2015 (Fieldhouse et al. 2016).

Tables organized in this way are not very easy to interpret, and it is difficult to tell which age groups are the most likely to vote, since the size of the different age groups varies. To help us make these comparisons more easily we therefore convert the cell counts into percentages, just as we did in Chapter 15. There are a number of ways in which we can calculate percentages, and it is important to be clear about whether we are calculating (or interpreting) row percentages or column percentages. Each type of percentage refers to something slightly different, so it is important not to get confused. Column percentages are the standard practice. These are read down the table in columns. These are reported in Table 16.4. So, for example, we can see that 56 per cent of 18–39-year-olds voted in the election. When interpreting the table, it is useful to draw comparisons between the different cells. For example, with reference to Table 16.4, we can see that turnout appears to increase with age. Whereas only 56 per cent of 18–39-year-olds reported having voted, 75 per cent of 40–59-year-olds and 86 per cent of the over 60s reported having done so. Reported turnout among the over 60s was therefore thirty percentage points higher than it was among the 18–39-year-olds. This difference appears to be quite substantial.

Row percentages are somewhat different. These are not reported in the table, but they can easily be calculated from the cell counts in Table 16.3. In this example, a row percentage would refer to the percentage of voters who fall into each age group. So, for example, we can see that 2,172 respondents reported that they voted in the election. Of these 2,172 respondents, 462 were aged 18–39, which equates to 21 per cent. By contrast, 35 per cent of voters were aged 40–59, and 43 per cent of voters were aged 60 or over. We have to be careful what substantive conclusions we can draw from these percentages. We can see that the largest group of voters are those who are aged 60 or over. This is because this age group is quite numerous in size (it accounts for 37 per cent of the sample) and also because this age groups is especially likely to vote (as we saw in Table 16.4). The over-60s are thus an important demographic for determining election outcomes, since if parties can do well among this group they stand to win a lot of votes. By contrast, the under 40s represent a much smaller pool of voters. This is perhaps one reason why parties often tend to target the votes of older people rather than younger people.

Table 16.4 Age and turnout in 2015 British General Election, column percentages

	18–39 years old	40–59 years old	60 years old and over	All
Did not vote	44	25	14	26
Voted	56	75	86	74
N	829	1,027	1,089	2,945

Source: Fieldhouse et al. (2016).

Table 16.5 Age and turnout in 2017 British General Election, column percentages

	18–39 years old	40–59 years old	60 years old and over	All
Did not vote	44	21	11	21
Voted	66	79	89	79
N	604	732	836	2,172

Source: British Election Study 2017 (Fieldhouse et al. 2018).

So what happened in 2017? And is there any evidence that turnout among young people was higher than previously? To answer these questions we can inspect Table 16.5, which shows the relationship between age and turnout in 2017. The first thing to notice is that reported turnout was about five percentage points higher than it had been in the previous election (79 per cent vs. 74 per cent). Once again we can see that older people were more likely to report having voted than younger people: 89 per cent of people aged 60 years or above said they voted compared to 66 per cent of 18–39-year-olds. Reported turnout among the over 60s was therefore twenty-three percentage points higher than it was among the 18–39-year-olds. This difference is still substantial though perhaps not quite as substantial as it was in 2015, when the age-gap in turnout was thirty percentage points. If we compare Tables 16.4 and 16.5 we can see that reported turnout among younger people in 2017 was 66 per cent, which is ten percentage points higher than it was in 2015. By contrast, reported turnout among older people in 2017 was 89 per cent, which is just three percentage points higher than it was in 2015. There is thus some evidence that turnout increased more among younger people than it did among older people.

However, we should treat these findings with some caution. First, we are dealing with estimates—and as we discussed in Chapter 15 with any estimate there is uncertainty. Second, we should be wary of drawing strong conclusions from small differences, and before agreeing that there was a youthquake we would want to know at least whether these differences were significant. And finally, we might want to reflect on the language we choose to describe the magnitude of this change—and think about what would be an appropriate adjective. Is it a small change? A big change?

Is there a relationship? Statistical significance for cross-tabulations

So far we have used cross-tabs as a descriptive tool to describe the observed pattern of association between two variables in our sample. In some instances, we may observe a clear pattern, where it is obvious (or at least appears to be so) that the distribution of responses varies by sub-group. But, in other instances, the observed pattern may be less clear-cut, and there may be only small differences between the sub-groups, or hardly any difference at all. Under these circumstances, it is not so obvious whether there is a 'real' relationship or not. Do the differences that we observe reflect real differences that exist within the population, or could they just be the result of sampling error?

To answer these questions we need to carry out tests of significance, based on something called chi square. As we discussed with reference to regression, there are different ways of carrying out these statistical tests and the test that we employ depends upon the type of

Research in Action 16.1 Was There a Youthquake in the 2017 UK General Election?

Following the 2017 UK general election, there was a debate about turnout. At 69 per cent, turnout was the highest that it had been for two decades. A number of potential reasons for this increase were put forward, but one popular narrative was that young people had turned out in their droves to support Jeremy Corbyn's Labour party. The press soon labelled this a 'Youthquake'. In the months that followed a number of different academic teams began to examine the issue more carefully. First the current BES team argued that the 'youthquake was a myth' (Prosser et al 2018). They defined the 'youth' as people aged between 18 and 24 years old, and showed that in 2017 turnout among this group of people was not significantly different to what it had been in 2015. Though turnout among the slightly older group of people aged 25 to 39 did increase, the authors say that this age group are too old to be considered youthful (even if by some definitions they may still be considered young). Importantly the BES team were able to cross-check whether the people who said they had voted in the survey actually had done so by inspecting the electoral register to see if they had been marked off as having voted. However, the BES teams' findings were challenged by another group of scholars, who argued that the sample size of 18- to 24-year-olds in the BES survey was too small to draw valid conclusions (Stewart et al. 2018). Using a different survey, with a larger sample, Sturgis and Jennings (2018) provided evidence that turnout among the 'youth' had increased. However, even though the survey they used had a larger sample size, it could not be used to measure turnout as accurately as could the BES survey, since it had not cross-checked whether the people who had said they had voted actually had done so. These examples illustrate that the different methodological choices and decisions that scholars make influence the conclusions they draw. Different surveys, with different sampling stratagies, sample size, and measures of turnout can produce slightly different results. There is not always a perfect data source and so we have to be aware of the limitations of the data we use and the conclusions that we can therefore draw. But even if the idea of a youthquake were contested, the weight of evidence across all three studies consistently shows that turnout among those aged 18 to 39 did increase. It might not have been a quake. And they may not have been particularly youthful. But they did turn out at the polls.

variable we are examining and the type of analysis that we are carrying out. However, all significance tests serve the same underlying purpose, which is to provide us with the information to make a decision based on probability as to whether or not we think the relationship we observe in our sample reflects a real relationship that exists in the population.

To illustrate these points, Table 16.6 shows two hypothetical examples. Example 1 depicts what it would look like if there were a perfect relationship between a person's social class (measured in terms of whether they self-identify as working class or middle class) and which party they reported voting for at the last election (just focusing on whether they voted for Labour or the Conservatives). By contrast, Example 2 depicts what it would look like if there were no relationship. From Example 1 we can see that all the working-class respondents voted for Labour and none voted for the Conservatives, whereas all the middle classes voted for the Conservatives and none voted for Labour. This is an example of perfect association. If we know somebody's class—that is, whether they are middle class or working class—we can perfectly predict which party they will vote for. By contrast, Example 2 depicts an example of (perfect) no association. We can see that exactly the same proportion of middle-class and working-class voters support each party. Knowing someone's class does not therefore help us at all in predicting which party they vote for. The conditional

Table 16.6 Association, hypothetical example, column percentages

	Example 1: perfect association		Example 2: no association	
	Working class (%)	Middle class (%)	Working class (%)	Middle class (%)
Labour	100	0	55	55
Conservative	0	100	45	45
Total	100	100	100	100

distributions of each class are exactly the same, and there is thus no association between the two variables. These two examples provide extreme scenarios. In practice, we are never likely to find the first example in the real world, but what we want to know is whether our sample is similar or not to the second example. If we think that it is similar, then we conclude that there is probably not a relationship between the two variables, but if we think that it is different from Example 2, then we can conclude that there probably is a relationship.

Let's introduce some real data. Before the 2017 UK election there was quite a lot of discussion about class. Theresa May's Conservative party campaigned on a platform to deliver Brexit and support 'ordinary working families' (the 'OWFs'). This was partly about trying to win back Conservative voters who had defected to UKIP, but also partly about trying to appeal to working-class voters in Labour-held seats who had voted for Brexit. By contrast, Jeremy Corbyn's Labour party campaigned on a much more radical left-wing platform, which opposed austerity and promised to work for 'the many, not the few'. Thus in different ways each of the main parties sought to appeal to working-class voters. So, who did the latter vote for?

Table 16.7 shows the column percentages for a cross-tab between class self-identification (whether voters identified as working class, middle class, or didn't identify with any class) and vote choice in the 2017 elections. We can see that Labour was more popular among the working class than it was among the middle class or among people who didn't identify with any class. Whereas 58 per cent of those who identified as working class reported voting for Labour, just 34 per cent of the middle class and 37 per cent of those with no class identity reported doing so. By contrast, the Conservative party was much more popular among the middle class (48 per cent of whom said they voted for the party) than the working class (where just 30 per cent said they voted for the party). Although the numbers for the other parties are somewhat smaller and so should be treated with some caution, we can see that the Liberal

Table 16.7 Class self-identification and vote, column percentages

	Class identification			
	Middle-class ID	Working-class ID	No class ID	All
Labour	34	58	37	44
Conservative	48	29	45	40
Liberal	10	4	8	7
Other	7	9	10	9
N	389	591	658	1,683

Source: British Election Study 2017 (Fieldhouse et al. 2018).

Democrats was more than twice as popular among the middle class (10 per cent of whom voted for the party) than it was among the working class (where just 4 per cent reported that they voted for the party). Class identity therefore appears to have been an important factor in understanding which party people supported. But are these differences significant?

As before, to establish whether our sample is different or not from what we would expect to see if there were no relationship between the two variables, we carry out a significance test. The method we use to test for significance in cross-tabs is called chi square (often denoted as χ^2). Chi square enables us to assess how likely it is that the sample association we observe could just be the result of sampling variation. Although the chi-square statistic that we use to test for significance in cross-tabs is different from the t test that we discussed with reference to bivariate analysis with interval data, the analytical principles are much the same and we go through the same basic series of steps:

- state null hypothesis;
- inspect statistic and interpret;
- accept or reject null hypothesis.

The null hypothesis is stated in exactly the same way as we discussed earlier. Using the example discussed before, the null hypothesis would be:

> H_0: In the population, there is no association between someone's class identity and which party they vote for.

If we can find sufficient evidence to reject the null hypothesis (at the 0.05 level), then we can infer that in the population there probably is an association between the two variables. The logic underpinning chi square is quite intuitive. It allows us to compare what we observe with what we would expect to observe if there was no association. We compare the table that we observe in our sample with a theoretical table of what we would expect to observe if there were no relationship between the variables (as shown in Example 2 of Table 16.6). To do this, we compare something called the observed frequencies and the expected frequencies. Starting with a table with just the cell frequencies, we compare the observed frequencies (O) that we observe in our sample, with the ones that would be expected under the null hypothesis of no association (E). The **chi-square test** is an excellent workhorse for testing whether the differences between O and E are statistically significant or not.

Table 16.8 shows the observed distribution of frequencies for the cross-tab of class by vote choice. The cell counts refer to the observed frequencies. So, for example, we can see that in our sample there were 341 people who identified as working class and voted Labour. The expected frequencies are calculated for the middle class and working class, assuming that the distribution of responses is exactly the same as for the sample overall. Since the total proportion of Labour voters in the sample is 0.438, we would therefore expect (if the null were true) 44 per cent of the working class to vote Labour. Therefore the expected frequency of working-class Labour voters would be 591 (the total number of working class in the sample) multiplied by 0.438 (the total proportion of Labour voters in the sample), which gives 259. This is somewhat lower than the observed frequencies for this cell, which was 341. This indicates that we have more working-class Labour voters in our sample than we would expect if there was no relationship between class and vote.

Table 16.8 Class self-identification and vote, observed frequencies

	Class identification			
	Middle-class ID	Working-class ID	No class ID	All
Labour	133	341	243	660
Conservative	187	174	299	717
Liberal	40	22	53	115
Other	29	54	63	146
N	389	591	658	1,683

Source: British Election Study 2017 (Fieldhouse et al. 2018).

BOX 16.3 How to Calculate Chi Square

The formula for calculating chi square is:

$$x^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

where:

f_o is observed frequencies; and
f_e is expected frequencies.

Chi square compares the difference between these expected and observed values for each cell in the table. The formula for calculating chi square is shown in Box 16.3. Although you will never need to, because computer programs will do it for you, it is very easy to calculate chi square. We can go through each of the steps with reference to Table 16.9. The column labelled f_o (observed frequencies) is what we observe in our sample. The column labelled f_e (expected frequencies) is the frequency distribution we would expect if the null hypothesis was true. The next column is simply the difference between these two values, and the column after that is the difference squared, so we get rid of the negative numbers (which would otherwise cancel themselves out when we add them all together). The final column divides the squared difference by the expected frequencies and then we simply add up these values for each cell and the total is chi square.

Clearly, a small chi square suggests that there probably isn't a relationship, since this means that the observed pattern is quite similar to the expected pattern assuming no association, and a large chi square suggests that there probably is a relationship between the two variables. But how do we decide what value of chi square constitutes large and what small? This is where probability theory comes to the rescue. Chi square has a known distribution, although the shape of this distribution varies according to how many cells we have in our table (known as degrees of freedom). We can thus compare our obtained chi-square statistic against the chi-square distribution for the appropriate number of degrees of freedom. If the probability of getting our obtained value of chi square is very small (typically below 0.05), then we can reject the null hypothesis. In this sense, the logic of significance testing

Table 16.9 Calculating chi square

	f_o	f_e	f_o-f_e	$(f_o$-$f_e)^2$	$(f_o$-$f_e)^2/f_e$
Middle-class Labour	133	170	-37	1,391	8
Working-class Labour	341	259	82	6,773	26
No Class ID Labour	243	288	-45	2,025	7
Middle-class Conservative	187	157	30	918	6
Working-class Conservative	174	238	-64	4,109	17
No Class ID Conservative	299	265	34	1,149	4
Middle-class Liberal	40	27	13	161	6
Working-class Liberal	22	42	-20	380	9
No Class ID Liberal	53	46	7	46	1
Middle-class Other party	29	35	-6	32	1
Working-class Other party	54	53	1	2	0
No Class ID Other party	63	59	4	19	0
∑	1,638	1,638	0	17,007	$x^2 = 86.2$

is the same as discussed previously with reference to the t test, only we are using a different instrument (chi square), which has a different type of distribution.

We should note, however, that chi-square tests provide an overall measure of whether or not there is a statistical association between our two variables. It thus provides a global test for the table as a whole. If we are interested in whether a particular cell within the table is significantly different from what we expect, then we need to look at something else. The adjusted residuals enable us to see in which particular cell of the table the Observed and Expected frequencies significantly differ (at the 5 per cent level). Adjusted residuals greater than plus or minus 1.96 thus indicate a significant difference.

Example 2: ethnic diversity and civil war

As before, we can put all these different steps of analysis together to examine a real-world problem. Research on the causes of civil war has received a great deal of attention in Politics and International Relations. The causes are hotly contested. Whereas Fearon and Laitin (2003) insist there is no link between ethnic heterogeneity and conflict, others find evidence that ethnic cleavages may increase the risk of conflict (Ellingsen 2000; Cederman and Girardin 2007). Part of the reason for the controversy is that different scholars use different data sets and different operational definitions of key variables. However, the fact that these different measures tend to produce different results should give pause for thought and raise concerns about the reliability and validity (see Chapter 6) of the different indicators.

To examine this association, we use data on civil war from the Uppsala/PRIO armed conflict data project (Gleditsch et al. 2002), supplemented with additional data from Gleditsch (2007). The data set provides information on 177 countries across every year from 1946 to 2002, detailing the incidence of armed conflict involving more than twenty-five casualties.

To measure the ethnic composition of the different countries, we follow Gleditsch (2007) and examine the share of the population which is not in the dominant ethnic group. The ethnic dispersion measure is therefore a hundred minus the percentage share of the dominant ethnic group, based on data provided by Vanhanen (2001), recoded into four equal categories ranging from societies which are very ethnically homogeneous (coded 1) to very ethnically diverse (coded 4).

Using this data, we examine the relationship between ethnic diversity and civil war. The results are presented in Table 16.10. Our null hypothesis is therefore that there is no association between the level of ethnic diversity in a country and whether or not it has experienced civil war. Our first task is to establish whether or not we can reject the null hypothesis. To do this, we inspect the chi-square statistics. The value of the Pearson chi square is 330 on 3 degrees of freedom. The probability of obtaining a chi square of this size or greater if the null hypothesis were true is less than 0.0005 (SPSS actually says it is 0.000, but it is not technically zero, so we just say $p < 0.0005$). This is clearly very small, and well below the conventional cut-off mark of 0.05. We can therefore confidently reject the null hypothesis and infer that there does appear to be a significant relationship between ethnic diversity and civil war.

Having established that we think a relationship does exist between the two variables, our next task is to describe the pattern of the relationship to see if it conforms to our expectations. To this end, we discuss the percentages in the table. First, we can see that in very homogeneous societies there does not appear to be a very high likelihood of civil war taking place. In very homogeneous societies, only 2.6 per cent of the cases experienced an incident of civil war. By contrast, in more ethnically mixed societies, the incidence of civil war appears to be higher. Among quite homogeneous societies, 20 per cent have experienced civil war; in quite diverse societies, 18 per cent have experienced civil war; and in very diverse societies 21 per cent have experienced civil war. These percentages are all substantially higher than what we observed for very homogeneous societies, lending some support to our initial hypothesis that civil war tends to be higher in ethnically diverse countries.

But interestingly, there does not appear to be a great deal of difference in levels of civil war between countries that are quite homogeneous (20 per cent) and countries that are quite diverse (18 per cent) or very diverse (21 per cent). This suggests that the relationship is not linear. It doesn't appear to matter how big the dominant ethnic group is, it just matters whether or not there is a dominant ethnic group.

Table 16.10 Civil war and ethnic diversity, column percentages

	Very homogeneous societies	Quite homogeneous societies	Quite diverse societies	Very diverse societies	Total
No civil war	97.4	80.5	81.9	79.3	85.0
Civil war	2.6	19.5	18.1	20.7	15.0
N	2,003	1,879	1,806	1,869	7,557

These results appear to confirm our expectations. Ethnic diversity, or more specifically, the presence of a dominant ethnic group, does appear to increase the likelihood of civil war. However, it is also important not to overinterpret these results. To be sure, diverse countries tend to experience more civil war than homogeneous countries, but even among diverse countries the vast majority (around 80 per cent) do not experience civil war, so the relationship is far from deterministic. We may therefore want to investigate further why some ethnically diverse countries descend into civil war whereas others do not. What other factors are important? How does ethnic morphology interact with economic conditions, institutional structures, and security relations with neighbouring countries? To answer these questions, we need to examine other factors, and it is this aim of developing multi-causal explanations that leads us to multivariate analysis, which we consider in Chapter 17.

Conclusions

This chapter has provided an introduction to statistical analysis using bivariate analysis. We have introduced the key concept of hypothesis testing, which enables us to see whether two variables are statistically associated and if so, how, using simple linear regression and cross-tabs. Bivariate analysis provides a useful starting point for examining how variables are related, and it is an important building block for developing explanations and causal statements about why things happen. But it does not tell us the full story. Inevitably, bivariate analysis leads us to ask more questions. In the real world, social and political phenomena are rarely mono-causal. If we are interested in why things happen, or why certain factors are influential in some conditions rather than others, the picture rapidly becomes more complicated and we are compelled to look for additional explanations, and this brings us to multivariate analysis, which we will discuss in Chapter 17.

Questions

1. Draw an arrow diagram depicting the relationship between two variables.
2. Write a brief note to accompany the arrow diagram, and state clearly which is the dependent variable and which is the independent variable, and why you think they may be related.
3. What is the null hypothesis? How is it tested?
4. Why should we treat the results from bivariate analysis with caution?
5. What is the difference between causality and association?
6. How do we know whether a relationship between two variables is causal or not?
7. In cross-tabs, when might it be useful to report the row percentages and when might it be useful to report the column percentages?
8. In simple linear regression, what is more substantively interesting: the magnitude of the slope coefficient or the value of R-square? What does each tell us?

Guide to Further Reading

Agresti, A. and B. Finlay (2009), *Statistical Methods for the Social Sciences*, 4th edition (Upper Saddle River, NJ: Pearson–Prentice Hall).
Covers the basics of statistical description and inference, as well as more advanced topics on regression methods, including multiple regression, ANOVA and repeated measures ANOVA, analysis of co-variance, logistic regression, and generalized linear models.

Carlson, J. and M. Hyde (2002), *Doing Empirical Political Research* (Boston, MA: Houghton Mifflin).
Provides an introduction to quantitative analysis and covers some more advanced topics, with lots of examples from political research.

de Vaus, D. (2002), *Surveys in Social Research* (London: Routledge).
A detailed and accessible introduction to quantitative methods, with easy-to-read chapters on how to carry out analysis and interpret the results.

Field, A. (2017), *Discovering Statistics Using SPSS*, 5th edition (London: Sage Publications).
Provides an easy-to-follow introduction to quantitative analysis and step-by-step instructions on how to use the SPSS statistical software package to do your own analysis.

Johnson, J. B., H. Reynolds, and J. Mycoff (2007), *Political Science Research Methods*, 6th edition (Washington, DC: CQ Press).
Provides an introduction to quantitative analysis and covers some more advanced topics, with lots of examples from political research.

Rowntree, D. (2000), *Statistics without Tears: An Introduction for Non-Mathematicians* (Harmonsdsworth: Penguin).
An easy-to-read introduction to statistical methods.

References

Cederman, L. and L. Girardin (2007), 'Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies', *American Political Science Review* 101: 173–85.

Ellingsen, T. (2000), 'Colorful Community or Ethnic Witches' Brew? Multiethnicity and Domestic Conflict During and After the Cold War', *Journal of Conflict Resolution* 44(2): 228–49.

Fearon, J. and D. Laitin (2003), 'Ethnicity, Insurgency, and Civil War', *American Political Science Review* 91(1): 75–90.

Fieldhouse, E., J. Green, G. Evans, H. Schmitt, C. van der Eijk, J. Mellon, and C. Prosser (2016), 'British Election Study, 2015: Face-to-Face Post-Election Survey' [data collection]. UK Data Service SN: 7972. http://dx.doi.org/10.5255/UKDA-SN-7972–1

Fieldhouse, E., Green, J., Evans, G., Schmitt, H., van der Eijk, C., Mellon, J., Prosser, C. (2018). British Election Study, 2017: Face-to-Face Post-Election Survey [data collection].

Ford, R. and M. J. Goodwin (2014), *Revolt on the Right: Explaining Support for the Radical Right in Britain* (Abingdon: Routledge).

Freedom House (2010), *Freedom in the World*. http://www.freedomhouse.org

Gleditsch, K. S. (2007), 'Transnational Dimensions of Civil War', *Journal of Peace Research* 44(3): 293–309.

Gleditsch, N., P. Wallensteen, M. Erikson, M. Sollenberg, and H. Strand (2002), 'Armed Conflict, 1945–99: A New Dataset', *Journal of Peace Research* 39(5): 615–37.

Marshall, M. G. and K. Jaggers (2010), *Polity IV Project: Political Regime Characteristics and Transitions, 1800–2010*, Center for Systemic Peace. http://www.systemicpeace.org/polity/polity4.htm

Norris, P. (2015), 'Democracy Cross-National Data', Release 4.0, Fall. https://www.pippanorris.com/data

Prosser, C., E. Fieldhouse, J. Green, J. Mellon, and G. Evans (2018), 'The Myth of the 2017 Youthquake Election'. https://www.britishelectionstudy.com/bes-impact/the-myth-of-the-2017-youthquake-election/#.XUbaaXt7nIX

Przeworski, A., M. Alvarez, J. Cheibub, and F. Limongi (2000), *Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990* (Cambridge: Cambridge University Press).

Stewart, M., H. Clarke, M. Goodwin, and P. Whiteley (2018), 'Yes, There Was a "Youthquake" in the 2017 Snap Election—And It Mattered'. https://www.newstatesman.com/politics/staggers/2018/02/yes-there-was-youthquake-2017-snap-election-and-it-mattered

Sturgis, P. and W. Jennings (2018), 'Why 2017 May Have Witnessed a Youthquake After All'. https://blogs.lse.ac.uk/politicsandpolicy/was-there-a-youthquake-after-all/

Svolik, M. (2008), 'Authoricarian Reversals and Democratic Consolidation', *American Political Science Review* 102(2): 153–68.

Vanhanen, T. (2001), 'Domestic Ethnic Conflict and Ethnic Nepotism: A Comparative Analysis', *Journal of Peace Research* 36(1): 55–73.

Varshney, A. (2002), *Ethnic Conflict and Civic Life: Hindus and Muslims in India* (New Haven, CT: Yale University Press).

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

A Guide to Multivariate Analysis

Chapter Summary

This chapter examines how we can develop and test multi-causal explanations of political phenomena. It builds on Chapter 16 by extending the principles of bivariate analysis to take into account more than one independent variable. In doing so, we examine two of the most widely used statistical analysis techniques in political research, ordinary least squares (OLS) regression and logistic regression. This chapter includes:

- causality and association;
- statistical control;
- spurious relationships;
- indirect causality;
- interaction effects;
- linear regression;
- logistic regression.

Introduction

This chapter examines multivariate analysis, which builds directly on many of the principles discussed in Chapter 16. Multivariate analysis involves examining the pattern of associations between more than one independent variable and the dependent variable. Political phenomena are rarely mono-causal. If we are interested in trying to explain 'why' things happen, then more often than not we need to develop explanations that take into account more than one explanatory factor. Bivariate analysis allows us to get an initial sense about the structure of the relationship between two variables. However, this may not accurately reflect the 'true' nature of the relationship for a variety of reasons. Undertaking multivariate analysis allows us to examine the impact of multiple factors on our dependent variable of interest, and to compare the explanatory power of rival hypotheses. It also allows us to examine whether the association between one independent variable and the dependent variable still holds up when we control for—or take into account—other factors, and to explore whether the impact of one variable also depends upon the level of another variable.

Investigating these issues helps us to develop fuller explanations for why different political phenomena occur, and provides a rigorous test of our explanations to see if they still stand up in the face of close scrutiny.

As we have seen in previous chapters, the way in which we carry out quantitative analysis depends upon the types of variables that we are interested in examining. The same applies to multivariate analysis. In this chapter we consider two different methods: ordinary least squares (OLS) regression (which is an extension of what we discussed in Chapter 16 for interval-level dependent variables) and logistic regression (which is appropriate when our dependent variable is categorical, specifically a dichotomy). These methods of analysis are among the most widely used quantitative techniques in political research, and understanding how they can be used and interpreted provides a strong foundation for critically evaluating much of the research that is published in leading academic journals.

The first part of this chapter examines the principles of multivariate analysis, and the different types of analytical question to which they can be applied. The second part examines multivariate analysis when our dependent variable is interval-level, using OLS regression that builds directly on the topics covered in Chapter 16. The third part then considers multivariate analysis when our dependent variable is categorical, using logistic regression, which builds on cross-tabulations (cross-tabs).

The principles of multivariate analysis: statistical control

Most social and political phenomena are multi-causal, which means that insights garnered from bivariate analysis, where only one causal variable is considered, are going to be partial at best. Multivariate analysis allows us to examine the impact of multiple causal factors on our dependent variable of interest in order to see which factors matter, or which matter most. It thus provides a useful tool for developing a more complete understanding of the topic. Multivariate analysis also allows us to have confidence in our results, and to assess whether or not they are robust. Do our results still stand up when we include other potentially important factors?

Bivariate analysis provides a good starting point for examining the structure of relationships between variables. It allows us to develop causal explanations and provide tentative answers to our research questions. But establishing causation is notoriously difficult, and we are probably never able to do it completely. At best, we therefore aim for 'robust association'. This requires three conditions. The first is statistical association. The second is appropriate time ordering. And the third is the elimination of other possibilities. From bivariate analysis, we can establish the first two conditions, but not the third. This last point is very important. To have confidence in our findings we have to subject them to rigorous testing. We have to be sceptical. If we observe a bivariate relationship, we always have to think about whether that relationship is real, or whether there are potentially other factors that could account for it. This is where multivariate analysis comes to the fore.

Suppose we are interested in exploring social trust. A great deal of research has examined the benefits of living in a community with high levels of social trust, and the factors that are associated with it. We might have three different hypotheses. Our first hypothesis might be that social trust tends to be lower in ethnically diverse communities, since people are more

likely to trust those who are like them. When there are many people from different ethnic backgrounds, the level of social trust in the community therefore tends to be lower, since, in the words of Robert Putnam (2007), people 'hunker down'. Our second hypothesis might be that social trust tends to be lower in economically deprived areas, since neighbourhoods with economic problems are more likely to be run down and lack the nice civic spaces that foster social trust, and so people may become socially withdrawn. And finally, our third hypothesis might be that social trust tends to be lower in areas with high levels of crime, since crime and the fear of crime stop people going out onto the street and interacting with each other, and therefore this fosters social isolation. From a bivariate analysis, we might find support for each of these hypotheses: that ethnically diverse areas tend to have lower levels of social trust than ethnically homogeneous areas; that economically deprived areas tend to have lower levels of social trust than economically prosperous areas; and that neighbourhoods with high levels of crime have lower levels of trust than neighbourhoods with low levels of crime.

A multivariate model allows us to examine the relative impact of all these variables on social trust when we examine them simultaneously. That is, when we control for—or take into account—a neighbourhood's level of economic deprivation and crime level, does its level of ethnic diversity still influence social trust? Another way of thinking about this is, supposing we are comparing two neighbourhoods with exactly the same level of economic deprivation and crime, but one neighbourhood has high levels of ethnic diversity and the other has low levels of ethnic diversity, would their respective levels of social trust differ? If their levels of social trust were different, then we could say that controlling for economic deprivation and crime, there is an association between ethnic diversity and trust. We can express the multivariate model that we want to test symbolically, using arrow diagrams, as shown in Figure 17.1.

Multivariate analysis allows us to compare the importance of these factors against each other, and to examine their independent impact on the dependent variable. For example, we might say that ethnic diversity and economic deprivation have no independent impact on levels of social trust, once we take into account levels of crime. Ethnically diverse neighbourhoods might tend to be relatively poor, and poor areas might have higher levels of crime, but it is only crime that is directly associated with levels of trust. That is, among neighbourhoods with high levels of crime, the level of trust does not vary by ethnic diversity or economic deprivation. Diverse, poor, safe communities have the same level of trust as homogeneous, rich, safe communities. The key thing is whether the communities are safe or not, not how rich or diverse they are. In this scenario, ethnic diversity and economic

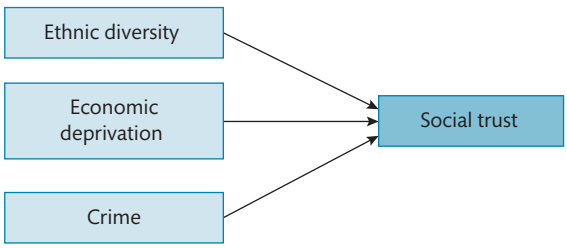


Figure 17.1 Multivariate model: predicting social trust

Research in Action 17.1 Does Diversity Erode Social Cohesion?

In recent years the public debate about social diversity and social cohesion has become increasingly politicized and often emotionally charged. In order to examine the link between community cohesion and diversity in Britain in a more systematic way, Natalia Letki (2008) examined the impact of racial diversity on particular dimension of social cohesion: social capital. Social capital refers to people's sense of community, their sense of belonging to a neighbourhood, caring about the people who live there, and believing that people who live there care about them. It therefore has a number of different dimensions. The results of Letki's analysis show that when the effects of diversity and neighbourhood deprivation on social capital are modelled simultaneously, diversity has a negative effect on only one attitudinal dimension of social capital to do with attitudes towards the neighbour. However, at the same time, the level of deprivation with a neighbourhood is a much more important factor undermining all types of interactions and positive attitudes among neighbours. These findings have important policy implications. Letki argues that the British government's efforts to de-emphasize socio-economic deprivation and focus on stimulating intercommunity 'unity from diversity' have been misplaced. Since deprivation is the major factor eroding community relations, attempts to re-invigorate social cohesion would be better placed focusing on this issue.

deprivation have no independent effect on social trust when we control for, or take into account, levels of crime. Diversity and social cohesion is explored further in Research in Action Box 17.1.

The idea of statistical control is one of the most important ideas in quantitative analysis. The driving force behind statistical control is to rule out theoretically plausible alternative explanations for the association between two variables to see if the association still stands up when we take into account—or control for—other variables. If we can rule out the alternative explanations, then this gives us more confidence that the way we have specified the relationship is the correct one. We can therefore have confidence that our results are robust, which ultimately is what we strive for. We want to be sure that they stand up. And the only way we can do this is to see if we can knock them down. This is a key strength of quantitative analysis. We are not interested in cherry-picking evidence to support our argument, but are interested in finding evidence that might contradict our argument. So if we think that the level of crime is really important for predicting a neighbourhood's level of trust, we have to think about all the reasons why it might not be. Is it crime, or the fear of crime, that drives social trust? What factors could account for the apparent association between levels of crime and social trust? If we can find other factors that can account for this statistical association, then we know our initial explanation is not very convincing. But if we can't, then we can have confidence that it might be reasonable.

Specifying different types of relationship

As the example in the previous section illustrates, when we move from bivariate analysis to multivariate analysis, we can sometimes uncover a different pattern of associations from those we initially observed. Multivariate analysis can help to knock down spurious relationships or even uncover real relationships that are hidden when we only examine two variables. But it

can also allow us to examine how variables are related to each other, distinguish between direct and indirect causal influences, and uncover how the impact of one variable is contingent on the presence (or absence) of another variable. We will discuss each of these in turn.

Spurious relationships

Spurious relationships occur when two variables appear to be associated, but there is in fact no causal relationship between the variables. The reason why the relationship is said to be spurious is that the independent variable is associated with both the probable cause and the outcome of a specific political phenomenon, but has no direct causal influence itself. Spurious relationships are most likely to occur when important explanatory factors are left out of the analysis. In these circumstances variables that are associated with the missing explanatory factors may look like they are associated with the outcome of interest, but this association is 'spurious' because it would disappear if the model were properly specified and the missing explanatory variables were included in the analysis. A real-life example of this is the classic study of student admissions to the University of California. Overall, it appeared that women had a less favourable chance of being accepted to the university than men, but this proved to be spurious once the subject they applied for was taken into account: women applied for more competitive subjects than men, but within each subject, their acceptance rates were at least as good as men's. Thus the original association between sex and acceptance was spurious, and subject can be regarded as a confounding variable.

We can further illustrate the logic of a spurious relationship with reference to the example discussed earlier of the relationship between a neighbourhood's ethnic diversity and level of social trust. We may observe a relationship between ethnic diversity and social trust (depicted in Figure 17.2). However, when we introduce a third variable, say economic deprivation, the significance of the relationship between diversity and trust disappears. That is, among neighbourhoods with the same level of economic deprivation, ethnic diversity has no impact on social trust. The reason for the apparent bivariate association, then, is that ethnic diversity is associated with an unmeasured third variable, which is economic deprivation (see Figure 17.3). Ethnically diverse neighbourhoods tend to be poorer than homogeneous areas, but, importantly, this is not regarded as a causal relationship.

Spurious relationships can also work the other way round, and there can be a spurious 'lack of association'. This is when a 'real relationship' between two variables is masked by a third variable. Obviously, to get a true idea of how variables are causally related we want to knock down any spurious relationships. This is why we have to think carefully about our **model specification**, and whether or not our explanation includes all the theoretically relevant causal factors that we can think of. Failure to control for a theoretically important variable can mean that we attribute undue causal importance to other variables. This is also known as omitted variable bias.

Figure 17.2 Bivariate relationship between diversity and trust

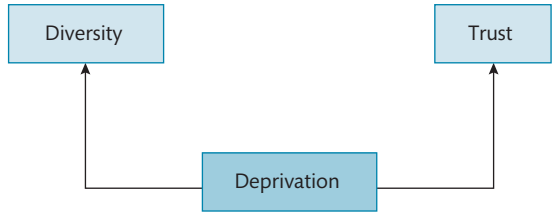


Figure 17.3 Deprivation as a confounding variable

Intervening variables and indirect causality

Intervening (or mediating) variables have a rather different role—they attempt to model the mechanisms through which, say, social class affects support for a particular political party. Possible intervening variables might be income or left–right values. For example, we might find from a bivariate analysis that there is a strong association between class and party support (see Figure 17.4), but that this association weakens (or even disappears) when we control for ideology. It might be tempting to think that the association between class and vote is spurious, as just discussed, and that class doesn't matter causally. But the key point that differentiates an indirect association from a spurious association is that we can plausibly postulate a causal sequence whereby different classes hold different ideological positions, which in turn affects one's propensity to vote for a particular party (see Figure 17.5). To be sure, these causal sequences are largely a matter of theory, but no theory could possibly suggest that, say, left–right attitudes determine one's class, although working-class people may well be disproportionately left wing.

Although indirect relationships and spurious relationships can look very similar on the surface, they have different theoretical implications. To distinguish between a spurious relationship and an indirect relationship we need to establish the temporal order of the variables and consider what is causally prior. So if we think diversity causes deprivation,

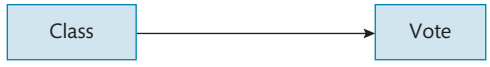


Figure 17.4 Bivariate relationship between class and vote

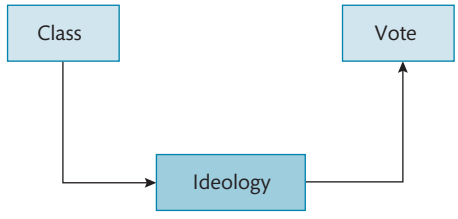


Figure 17.5 Ideology as an intervening variable

we might think that there is an indirect relationship between ethnic diversity and trust, and that economic deprivation is therefore a mediating variable. However, if we think that economic deprivation is causally prior to ethnic diversity, and that poor areas tend to be more ethnically diverse because immigrants are more likely only to be able to afford to live in poor areas (or are more likely to be allocated housing in poor areas), then we would say that the apparent relationship between ethnic diversity and trust is spurious.

Moderator variables and interaction effects

So far we have assumed that the introduction of a third variable works in much the same way for everyone. That is, we are dealing with uniform effects. But it is also often theoretically plausible that some factors have greater influence on some groups of people (or cases) than others. An **interaction effect** is when the impact of an independent variable is conditional on the presence of another independent variable, or moderator variable. So, for example, suppose we want to investigate the relationship between someone's policy preferences and their vote choice. We might think that those who hold left-wing preferences are more likely to vote Labour and those who hold right-wing preferences are more likely to vote Conservative. There is thus a bivariate relationship between ideology and vote (see Figure 17.6). But we might also think that the form of this relationship varies across different groups. In particular, we might think that the relationship between ideology and vote choice is stronger for people who have high levels of political knowledge (since perhaps they are more aware of the ideological differences between the parties), and weaker among people with low levels of political knowledge (since perhaps they are less aware of the ideological differences between the parties). The impact of ideology on vote choice is therefore contingent upon political knowledge, which can be regarded as a moderator variable. We can summarize this type of relationship symbolically, as shown in Figure 17.7, which shows that it is the combination of ideology and knowledge together (the interaction) that is important for predicting vote choice.

We now turn to how these different types of relationships can be examined in practice, and discuss the principles of multivariate analysis using OLS regression and logistic regression.

Figure 17.6 Bivariate relationship between ideology and vote

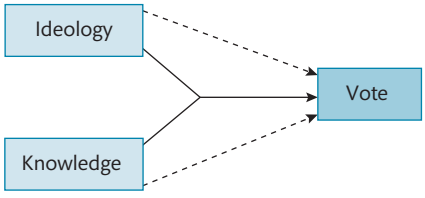


Figure 17.7 Knowledge as a moderating variable

Multivariate analysis using OLS regression

When our dependent variable (the thing that we are trying to explain) is interval-level (or ordinal if the variable approximates a normal distribution), the appropriate method of analysis is OLS regression, as discussed in Chapter 16. Multivariate analysis using OLS is just a simple extension of the simple linear regression model, and doesn't involve any new major concepts or techniques. We specify and estimate the model in much the same way (only with more independent variables) and interpret the results with reference to the significance, direction, and magnitude of the different parameter estimates. In this section we examine how we can specify a multivariate model and interpret the results from multiple regression, with reference to the significance and direction of the parameter estimates, and the fit with the data. We also show how the results can be presented, using figures to plot the predicted values of the dependent variable for different combinations of the independent variables.

Example 1: democracy

To illustrate the different steps we need to take to estimate and interpret a multivariate model using OLS regression, we build on the example we considered in Chapter 16 to do with the factors that are associated with a country's level of democracy. A great deal of research has examined the factors that are associated with the level of democracy in a country, and there are three common explanations for why some countries are more democratic than others. These refer to economic development, ethnic diversity, and state formation. The classic (1959) study by Seymour Lipset suggested that rich countries tend to be more democratic than poorer ones since economic development generates the wealth, social conditions, and cultural values which are conducive to liberal democracy. More recently, it has also been argued that ethnically diverse countries tend to be less democratic than more homogeneous countries, since ethnic differences can divide society and make compromise

BOX 17.1 Dummy variables

In Chapter 16 we examined the association between two interval-level variables, but we can also use OLS when one (or more) of our independent variables is measured at the nominal level. To do this, though, it is necessary to transform the nominal variable into what is known as a dummy variable. These are binary variables. Even if the nominal variable contains more than two categories, we can break it down into a series of 'dummies'. An example of a nominal-level variable is the continent or geographic region in which a country is located. The variable may contain six categories: Europe, Africa, Asia, North America, South America, and Oceania. To estimate the association between geographic region and, say, a country's level of democracy using OLS regression, we need to recode the regional variable into six different dichotomous variables, where each variable is coded 1 for the region of interest and 0 for other countries. These variables (apart from one) can then be entered into the regression equation. The variable that is not entered into the equation constitutes the reference category (e.g. 'Europe'). The coefficients for each of the other variables can then be interpreted as describing how much more (or less) democratic the countries in that regional are compared to the reference category (countries in Europe).

and consensus more difficult. Finally, it has also been argued that countries that have only recently gained independence may be less democratic than countries that have been independent a long time, since the institutions of the state may be weaker and less organizationally developed.

Before carrying out a multivariate analysis, it is always a good idea to look at how the independent variables are related to each other. The assumption is that they are independent—that is, they measure different things, both conceptually and empirically. If this assumption is violated, then there could be a problem of **multicolinearity**. This occurs when two independent variables are so highly correlated that they do not meaningfully measure different things. For example, if we had two independent variables, such as gross domestic product (GDP) per capita and the percentage of the population with a high-speed internet connection at home, we might find that the two variables are very highly correlated. That is, rich countries all have a high degree of internet penetration, and all countries with high levels of internet connection are rich. Under these circumstances, it becomes almost impossible to estimate the independent effect of each factor (since they are not really empirically different from one another). It is therefore advisable to either drop one of the variables from the analysis, or to combine the two variables together to form one composite measure, say of economic and technological development. There is no fixed cut-off point for how correlated is too correlated, but generally when the correlation coefficient between two independent variables is more than plus or minus 0.7, there is a problem.

Table 17.1 reports the Pearson's correlation coefficient (Pearson's r; see Chapter 16) between each pair of variables used in the analysis. The asterisk denotes whether the bivariate relationship is significant at the 0.05 level. We can see that, at the bivariate level, each of the independent variables is significantly associated with the level of democracy. The correlation coefficient between GDP per capita and democracy is 0.39, which is significant and positive, indicating that richer countries tend to be *more* democratic than poorer countries. The correlation coefficient between ethnic diversity and democracy is –0.27, which is also significant, but this time negative, indicating that ethnically diverse countries tend to be *less* democratic than ethnically homogeneous countries. Third, the correlation coefficient between year of independence and democracy is –0.24, which is significant and negative, indicating that countries which have become independent recently are *less* democratic than countries which became independent a long time ago.

Turning to the correlations between the independent variables, we can see that there is a significant negative correlation between GDP per capita and ethnic diversity

Table 17.1 Factors associated with democracy, bivariate correlation, Pearson's r

	GDP per capita	Ethnic diversity	Year of independence	Polity IV democracy
GDP per capita	–	–0.46*	–0.33*	0.39*
Ethnic diversity	–0.46*	–	0.14	–0.27*
Year of independence	–0.33*	0.14	–	–0.24*
Polity IV democracy	0.39*	–0.27*	–0.24*	–

Notes: *Significant at $p = 0.05$ level.

(Pearson's r = –0.46). This indicates that rich countries tend to be *less* ethnically diverse than poor countries. Similarly, there is a significant negative relationship between GDP per capita and year of independence (Pearson's r = –0.33), indicating that richer countries tend to have achieved independence earlier than poorer countries. Finally, there is not a significant correlation between year of independence and ethnic diversity (Pearson's r = 0.14). Importantly, though, none of these correlations is sufficiently large to indicate that there may be a problem with multicolinearity. Note that, with correlation (unlike regression), the relationship is symmetrical. This means that its value does not depend upon which variable is considered the independent variable and which is considered the dependent variable. Thus, we can see that the correlation between GDP per capita and ethnic diversity is exactly the same as the correlation between ethnic diversity and GDP per capita.

From the bivariate analysis reported above, we can see that there is a significant negative association between ethnic diversity and level of democracy. That is, more ethnically diverse countries tend to have lower levels of democracy. But how much can we infer bivariate association from this? Would we be justified in concluding that ethnic diversity inhibits democracy in some way? We may have theoretical arguments for why it might, and the data might be consistent with these expectations, but to have confidence in the result, we need to rule out the possibility that other factors are responsible for the apparent correlation.

For example, we may suspect that there is in fact no direct causal relationship between ethnic diversity and democracy. Rather, it may be the case that ethnically diverse countries tend to be rather poorer than more homogeneous countries (as we have also seen from the correlations reported above) and that it is this correlation that is driving the apparent association between ethnic diversity and democracy. Accordingly, when we control for economic development, we may find that the association between ethnic diversity and democracy disappears. To examine these possibilities, we can use OLS regression to examine the impact on democracy of all three independent variables. We can specify the model symbolically (as shown in Figure 17.8) and algebraically as follows:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3$$

where:

X_1 refers to ethnic diversity on a fractionalization scale (ranging from 0 to 1, where 0 is homogeneous and 1 is diverse);
X_2 refers to period of independence (where 0 = Pre-Second World War and 1 = Post-First World War);
X_3 refers to GDP per capita in US$1,000s.

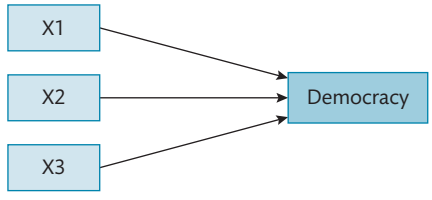


Figure 17.8 Multivariate model predicting democracy

Table 17.2 Predicting the level of democracy: OLS regression, unstandardized beta coefficients

	Model 1		Model 2		Model 3	
	Beta coefficient	Standard error	Beta coefficient	Standard error	Beta coefficient	Standard error
Constant	6.90*	1.10	12.49*	1.92	3.06	4.50
Diversity	−7.67*	2.11	−5.31*	2.14	−3.31	2.28
Independence			−3.99*	1.15	−2.91*	1.22
GDP					0.89*	0.39
R-square	0.09		0.16		0.19	
Adjusted R-square	0.08		0.15		0.18	

Notes: *significant at p<0.05 level, N = 145.
Source: Pippa Norris (2015).

Now, rather than just having one slope coefficient b to estimate (as we did with simple linear regression in Chapter 16), we have three slope coefficients to estimate. These are labelled b_1, b_2, etc. These coefficients are known as **partial regression coefficients**. They are partial because their magnitude depends upon the other independent variables in the equation. That is, they indicate how much the dependent variable changes for a one-unit increase in the independent variable, holding all the other variables in the model constant. To give a clear idea of how the independent variables are related to each other and to democracy, we will not estimate the full model all in one go, but will gradually build it up by introducing each independent variable into the model one at a time. We thus estimate three models. The first model just contains ethnic diversity as an independent variable; the second model contains ethnic diversity and period when gained independence; and the third model contains all three independent variables of ethnic diversity, period when gained independence, and GDP per capita. The results are presented in Table 17.2.

Interpreting tables using OLS regression is relatively straightforward. The key things to look for are whether the independent variables are significant or not (usually at the 0.05 level) and whether the impact of the variable is in the expected direction (positive or negative). We can also inspect the magnitude of the coefficient to see if it is substantively interesting.

Model 1 reports the effect of ethnic diversity when it is the only explanatory variable in the model. We can see that the coefficient is significant at the 0.05 level (indicated by an asterisk) and the sign of the coefficient is negative, which is in the expected direction. That is, more ethnically diverse countries tend to be less democratic. The magnitude of the coefficient is $b = -7.67$. This tells us that, for every one point increase on the fractionalization scale (which ranges from 0 to 1), the expected level of democracy in a country declines by 7.67 points on the Polity IV scale (which ranges from 0 to 10). Taken on its own like this, the magnitude of the coefficient refers to the total effect of ethnic diversity on democracy. When we include additional variables, part of this effect may be accounted for by other factors, and so what we will be left with is the partial effect—that is, the impact of ethnic diversity, controlling for the impact of the other variables in the model.

The last rows of the table report the R-square (see Chapter 16) and the adjusted R-square. As we add extra independent variables to the model, the adjusted R-square will only increase if the new term improves the model more than would be expected by chance. The adjusted R-square is therefore usually somewhat lower than the R-square. The interpretation of the adjusted R-square is also somewhat different from that of the R-square, and no longer refers to the 'explained' variance, since this has been adjusted by the number of variables in the model. It therefore takes into account the parsimony of the model, and is higher for a model which explains a given amount of variation in the dependent variable using only a few independent variables than for one that explains the same amount of variation but uses more independent variables. It is therefore often a good idea to report both measures. The R-square of 0.09 indicates that ethnic diversity explains 9 per cent of the variation in levels of democracy.

In Model 2, we now have two independent variables—ethnic fractionalization and whether the country achieved independence after the Second World War or not. We can see that both terms are significant and in the expected direction. That is, ethnically diverse countries still tend to be less democratic than ethnically homogeneous countries, even when we take into account when they achieved independence. So the relationship between ethnic diversity and democracy holds both for countries that achieved independence before the Second World War and for countries that achieved independence after the Second World War. In addition, we can see that countries that achieved independence after the Second World War tend to be less democratic than countries which achieved independence before it ($b = -3.99$), controlling for ethnic diversity.

However, interestingly, we also note some change in the magnitude of the coefficient for the ethnic diversity term. It decreases from $b = -7.67$ when it is the only variable in the model, to $b = -5.31$ when we control for the effect of independence. This tells us that part of the apparent association between ethnic diversity and democracy is really a function of independence, and when we take into account when the country achieved independence, ethnic diversity does not appear to be quite so important. This is because those countries which achieved independence after the Second World War tend to be more ethnically diverse than those countries which achieved independence before.

Finally, in Model 3, we examine the relative impact of all three independent variables simultaneously. This is our final model. We now see that, controlling for when a country achieved independence and its level of GDP per capita, the ethnic diversity of a country has no significant impact on its level of democracy. By contrast, the effects of independence and GDP per capita are both significant and in the expected direction. Countries which became independent after the Second World War tend to be less democratic than countries which became independent before it (controlling for the other variables in the model), and rich countries tend to be more democratic than poor countries (controlling for the other variables in the model). Controlling for GDP therefore completely wipes out the effect of ethnic diversity. When we know how rich a country is and when it achieved independence, it doesn't make any difference to our prediction of how democratic a country is if we know how ethnically diverse it is.

How then do we account for the apparent association between ethnic diversity and democracy at the bivariate level? Well, on the face of it, this appears to be a case of spurious correlation. Although ethnically diverse countries are less democratic than ethnically homogeneous ones, this isn't really anything to do with their level of ethnic diversity, it is more to do with their level of economic development. It just so happens that ethnically diverse countries tend to be poorer than ethnically homogeneous countries, and it is this correlation that drives the

apparent association between ethnic diversity and democracy. Of course, if we could think of some theoretically plausible reason for why the ethnic heterogeneity of a country might influence its level of economic development, then we could make a case for ethnic diversity having an indirect effect on levels of democracy. But, in the absence of this theory (which appears unlikely), the association is probably more likely to be spurious.

Measures of fit

The R-square for the final model is 0.19. We can interpret this in a number of ways. First, we can say that our model explains 19 per cent of the variation in levels of democracy between the countries. Or, to put it another way, we can say that if we are interested in predicting a country's level of democracy, we will make 19 per cent fewer errors if we know their scores on the independent variables than if we just guess based on the mean. We might not think this is very good. To be sure, there are numerous other factors that we could include in our model that might improve its predictive power. But at the same time, the importance of R-square should not be exaggerated. R-square is rarely the most important part of the model's results. We may be particularly interested in R-square if the regression model is fitted solely for the purpose of predicting future observations of the dependent variable. More often, however, we are at least as, or more, interested in examining the nature and strength of the associations between the dependent variable and the explanatory variables and testing specific hypotheses, in which case the regression coefficients are the main parameters of interest.

As discussed previously, we can interpret the regression coefficients in terms of their significance and their sign, and whether or not they support our theoretical expectations. Entering the variables in a step-wise fashion also allows us to inspect how the patterns of association change when we control for additional variables, which can shed light on direct and indirect causal mechanisms, as well as uncovering spurious associations. But we may also be interested in making some statement about which variable appears to be the most important. What factor has the biggest impact on the dependent variable? What matters most for a country's level of democracy? Its economic development? Its length of time as an independent state? Its ethnic diversity?

Interpreting the magnitude of the coefficients

We cannot make direct comparisons between the regression coefficients for the different independent variables because each variable is measured on a different scale. Since the regression coefficient tells us how much the dependent variable changes for each unit increase in the independent variable, the magnitude of the coefficients are all based on different scales, making direct comparison difficult. For example, independence is measured as a dummy variable (Box 17.1), GDP per capita is measured in thousands of US dollars, and ethnic diversity is measured in terms of fractionalization. We cannot say that each of these units is equivalent, and so direct comparison between the sizes of the coefficients is essentially meaningless.

However, it is often useful to be able to discuss the relative impact of different variables. Even if we cannot directly compare which variables 'matter the most', there are a number of techniques that we can use to try and do this indirectly. We could just enter the variables in the model one by one in step-wise fashion and look at which variables lead to the largest increase in R-square. However, the problem with this approach is that how much each

variable appears to 'explain' will partly depend upon the order in which it is entered into the model, and variables which are entered early can more easily explain a lot of variation than variables entered towards the end.

Alternatively, we could just compare how much difference it makes to the dependent variable when we increase each independent variable from its minimum value to its maximum value. So, for example, controlling for all the other factors in the model, what is the expected difference in levels of democracy between the poorest country and the richest country compared with the expected difference between the most ethnically homogeneous country and the most ethnically diverse country? The problem with this approach, though, is similar to the problems associated with looking at the range, which we discussed in Chapter 15 with reference to measures of variation. Accordingly, since how much each variable appears to matter depends upon the extreme values of the distribution, this approach is very sensitive to outliers. One or both of these values might be atypical, giving a misleading impression of how much the independent variable really matters.

One way we can get round this problem is to base our comparisons on the standard deviation instead (see Chapter 15). From probability theory we know that a given proportion of cases fall within one standard deviation of the mean. If we then compare the amount our dependent variable changes in relation to a one-standard-deviation increase for each of our independent variables, then we are comparing the change in the dependent variable in relation to broadly equivalent changes in the distribution of our explanatory variables. This 'broad equivalence' is often discussed in terms of what are known as standardized beta coefficients, or standardized regression coefficients, often denoted (capital) *B*. Most statistical software programs routinely include these measures as part of their output. 'Standardizing' a variable involves subtracting its mean from each individual value, and dividing by the standard deviation. The metric for a standardized regression coefficient is therefore standard deviation units. This means we can draw comparisons between the change in *Y* for a broadly equivalent change in the distribution of *X* (measured in standard deviation units).

The final model is reproduced in Table 17.3, with information on the unstandardized beta coefficient, the standard error, the standardized beta coefficient, the t value, and the

Table 17.3 Predicting the level of democracy: OLS regression, parameter estimates

	Beta coefficient	Standard error	Standardized beta coefficient	t value	p value
Constant	3.06	4.50			
Ethnic diversity	−3.31	2.28	−0.13	−1.45	0.15
Independence	−2.91*	1.22	−0.21	−2.38	0.02
GDP per capita	0.89*	0.39	0.22	2.31	0.02
N	145				
R-square	0.19				
Adjusted R-square	0.18				

Notes: *significant at 0.05 level.
Source: Pippa Norris (2015).

p value for each of the independent variables. We can see that the only terms for which the p value is below 0.05 are independence and GDP per capita. (Recall that the p value provides the probability of obtaining a test statistic of the observed size or greater if the null hypothesis is true.) From the standardized beta coefficients, we can get an idea of the relative importance of each variable in explaining variation in levels of democracy. Given that the ethnic diversity term is not significant, there is no need to discuss it further. However, we can see that the standardized regression coefficient for the independence term (B = −0.21) is of almost the same magnitude as the coefficient for the GDP per capita term (B = 0.22). Both terms therefore appear to have much the same relative impact.

Presenting the results

Finally, it is often helpful to present results graphically, as this can illustrate the impact of certain key variables, and provide a clear visual representation of the results for people who may shy away from looking at tables filled with numbers and asterisks. We can do this in much the same way as we did for bivariate regression (see Chapter 16), and use the regression equation to predict values of Y, depicted algebraically as follows:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3.$$

From our analysis, we have estimates for each of the different parameters, and so can substitute the values of each unstandardized regression coefficient into the equation as follows:

$$Y = 3.1 - 3.3X_1 - 2.9X_2 + 0.9X_3.$$

The last step is simply to plug in different values for X. We can then predict values of Y for countries that gained independence after the Second World War (when $X_2 = 1$) at different levels of GDP per capita, and do the same for countries that experienced independence before the Second World War (when $X_2 = 0$) in the same way. The predicted values of Y can then be illustrated graphically, as shown in Figure 17.9.

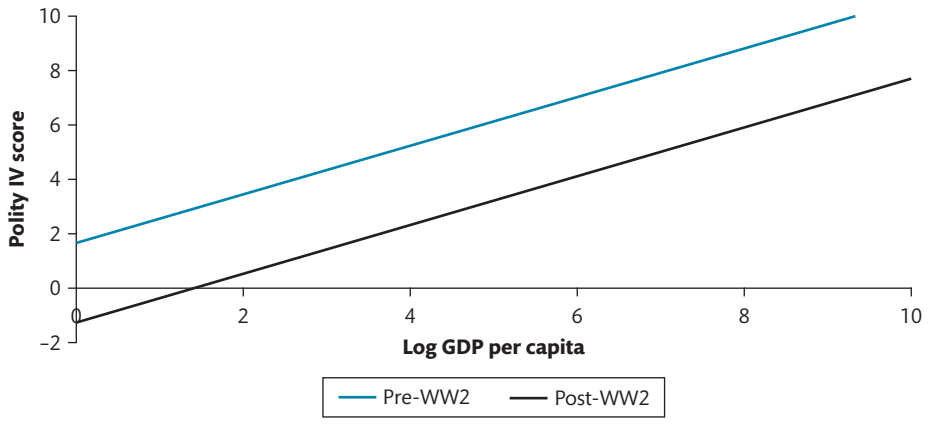


Figure 17.9 Estimated impact of GDP and period of independence on level of democracy

Research in Action 17.2 The Decline of Social Democracy: When Can the Mainstream Left Succeed?

In recent years, many mainstream social democratic parties have suffered at the polls. At the same time there has been a surge in support for populist radical-right parties in Europe. This has revived a decades-old debate about the crisis of the centre-left and social democracy, and the factors that may be associated with their decline. In a recent article Tarik Abou-Chadi and Markus Wagner (2019) set out to explore the conditions under which centre-left parties suffer set-backs, and to identify the strategies that they can employ to succeed at the polls. Abou-Chadi and Wagner argue that when centre-left parties adopt economically left-wing policies which prioritize investment, coupled with socially liberal policies that support immigration they perform more strongly in elections. To test this hypothesis they collect data on the electoral performance of social democratic parties in twenty-two advanced industrial democracies since 1975. Using OLS regression they then estimate the impact of the different policy positions that the social democratic parties adopt on their electoral performance. In addition they also control for a number of other factors that might affect support for left-wing parties, such as the level unemployment, GDP growth, and trade union influence. The results broadly support their hypotheses. However, Abou-Chadi and Wagner caution that this should not imply that there is a simple election winning formula that social democratic parties can follow. Although the combination of investment and liberal policies can boost support for social democratic parties by attracting the votes of the middle class, it can also harm the party by making it lose the votes of the working class, who are less supportive of immigration.

Multivariate analysis using logistic regression

So far we have considered using multivariate analysis for when our dependent variable is measured at the interval level. However, many political phenomena that we are interested in are not measured in this way. For example, those doing research on voting behaviour (see Research in Action Box 17.2) are often interested in which party people vote for (e.g. Labour, Conservative, Liberal Democrats, or other), or even whether people vote at all in the first place (i.e. vote or not vote). Similarly, we may be interested in whether or not a country has experienced civil war or ethnic conflict or terrorism in a given period of time (i.e. measured in terms of whether they have or have not). Indeed, research on democracy often treats democracy as a categorical variable, and distinguishes between whether a country is democratic or not, or even whether a country is a liberal democracy, electoral democracy, or some other type of 'democracy with adjective'.

When our dependent variable (the thing that we are trying to explain) is a binary variable, the appropriate multivariate technique to use is logistic regression. Our independent variables can be measured at either the nominal, ordinal, or interval level. This builds on some of the ideas and techniques we discussed with reference to cross-tabs in Chapter 16, and is similar in some respects to the principle of OLS regression discussed in the previous section. We use logistic regression to examine the patterns of association between our dependent variable and different independent variables. We first look to see whether the association is significant; then we look at the direction of the association—that is, whether it is positive or negative; and then lastly, we can make some substantive point about the magnitude of the effect—that is, whether it is large or small.

But logistic regression also involves the introduction of some new ideas. The main analytical difference with OLS regression is that the logistic regression coefficients have a different numerical interpretation. This is because they refer to odds and probabilities of change in the dependent variable, rather than simple increases or decreases in the dependent variable. Whereas the coefficient in OLS refers to the *amount of probable change* in the dependent variable—that is, how much we expect the dependent variable to change for a one-unit increase in the independent variable, the coefficient in logistic regression refers to the *probability of change* in the dependent variable.

With OLS regression, the coefficients refer to the impact that a one-unit increase of the independent variable has on the value of the dependent variable. With logistic regression, the coefficients do not refer to the value of the dependent variable—because it can only be one of two values—either zero or one. Instead, the coefficient refers to the **odds** or probability of the dependent variable taking one value rather than the other. So, for example, if our dependent variable is whether a country experienced civil war (coded 1) or not (coded 0), the coefficient for an interval-level variable tells us whether the probability of a country experiencing this outcome (or the odds of them having a civil war) increase or decrease for every unit change in the independent variable. By contrast, the coefficient for a categorical-level variable tells us whether the odds of experiencing civil war are higher or lower for the category in question in relation to the reference category.

If it were simply a question of dealing with changes in the probability of the dependent variable taking one value over the other, then the output from logistic regression would be quite intuitive and easy to understand. But this is not the case. For technical reasons, logistic regression involves transforming probabilities into something called **odds ratios** and **log odds ratios**. The statistics behind how these are calculated, and how the parameters are estimated, are quite complicated, and they are beyond the scope of this chapter (but see Agresti 2002 for a detailed introduction). We therefore focus on how they can be interpreted, which is far more straightforward.

The workhorses of logistic regression are odds and odds ratios. The idea behind odds is fairly uncomplicated, and it is closely linked to probabilities and percentages. So, for example, according to the 2017 British election survey, turnout was 79 per cent (see Table 17.4). We can convert this percentage into a probability that someone plucked at random would have voted—this probability is expressed as 0.79. We can phrase this in a different way and ask, what are the odds that a person plucked at random voted? The odds are calculated as the probability of having voted against the probability of not having done so. This can be expressed as: $p(v)/p(1\text{-}v) = 0.79/0.21 = 3.76{:}1$. The odds of voting of 3.76:1 tell us that for every one person who did not vote, there were nearly four people who did vote.

Table 17.4 Age and turnout in 2017 British General Election, column percentages

	18–39 years old	40–59 years old	60 years old and over	All
Did not vote	44	21	11	21
Voted	66	79	89	79
N	604	732	836	2,172

Source: Fieldhouse et al. (2018).

Table 17.5 Voting by age group, odds ratios

	18–39 yrs	40–59 yrs	60+ yrs
Voted	(REF)	2.5	5.4

Source: Fieldhouse et al. (2018).

Odds ratios extend this type of analysis by simply comparing the odds of two different groups. There are many different ways in which we can do this, but one comparison that we might be interested in is the odds of young people voting compared to older people. From Table 17.4 we can calculate the odds of voting for each of these age groups in the same way as above. The odds of voting among young people can therefore be calculated as: 0.66/0.44 = 1.5. And the odds of voting among older people are: 0.89/0.11 = 8.1. From these odds we can see that the odds of voting are higher among older people (8.1:1) than they are among young people (1.5:1). The odds ratio simply tells us how much higher (or lower) the first set of odds are compared to the last set (our reference category). So, the ratio between the odds of older people voting compared to the odds of young people voting (our reference category) can be calculated as: 8.1/1.5 = 5.4 (see Table 17.5). This odds ratio tells us that the odds of voting among older people are about five times higher than the odds of voting among young people.

Odds ratios can take values between 0 and infinity, but they are always positive numbers. If the odds ratio is 1, then the odds of voting are the same for both groups. If the odds ratio is less than 1, then the odds of voting among older people is less than the odds of voting among young people (which is our reference category). And if the odds ratio is greater than 1, the odds of voting for older people are greater than the odds of voting for young people. Odds ratios are therefore not symmetrical. For example, if our reference category is young people, then an odds ratio of 5 tells us that the odds of voting are five times higher among older people than among young people. However, if our reference is older people, then an odds ratio of 0.20 tells us that the odds of voting are five times lower among young people than among older people.

Odds ratios are a very useful way of describing the association between two variables. They also form the basis for interpreting logistic regression coefficients. Although the output from most statistical software packages does report the odds ratios, something called the log odds ratios are usually reported when writing up the results in tables. These are based on the odds ratios, but are transformed using something called the log function, so they are centred on zero. Log odds with positive values therefore indicate that the likelihood of voting is higher among the group in question than it is among the reference category, whereas log odds with negative values indicate that the likelihood of voting is lower among the group in question than it is among the reference category. It is therefore important always to be clear whether you are reporting (or interpreting) log odds ratios or odds ratios.

Although the interpretation of these coefficients might lack the intuitive simplicity of the OLS regression coefficients, logistic regression is really no more complicated to use or understand than OLS regression. The principles of statistical control are much the same, as are the principles of hypothesis testing. As with OLS regression, we are primarily interested in whether our independent variable has a significant association with the dependent variable (at the 0.05 level or above) and, if so, whether the direction of the relationship is positive

(indicated by positive log odds) or negative (indicated by negative log odds). We can then make some substantive interpretation about the magnitude of the effect, but this is often most easily done by transforming the log odds back into probabilities (which we will do later in this chapter).

Example 2: turnout

In recent years, there has been a great deal of concern about declining levels of turnout in Britain and many other advanced democracies. In particular there has been concern about the low participation rates of young people. Various different explanations have been proposed in order to try and explain why people vote or do not vote, and how this relates to their age. In this section, we focus on three different sets of factors that are widely thought to be associated with whether or not people vote. The first set of factors focuses on the social characteristics of voters, such as their age, sex, and education. There is a large body of research that examines participatory equality, and the extent to which people from different social backgrounds participate in politics. It is of descriptive interest to know whether participation rates vary by social background, but it is also of analytical interest to know whether those people with high levels of 'civic resources' such as education are more likely to vote, and whether people acquire the 'habit' of voting as they get older. The second factor relates to how attached people feel to political parties. If people have a strong attachment to a particular political party and feel that it 'represents' them in some way, then we may expect that they will be more likely to vote. Third, we can also look at factors related to the structure of political competition, such as whether the election is competitive and the race is expected to be close, and whether there is much policy difference at stake in terms of how different the parties are viewed as being from each other.

By examining which of these factors influence turnout, and which factors influence turnout the most, we can try and get a sense of what is most important. Is turnout something about voters? Something about parties? Or something about competition? How do these different factors relate to and interact with each other? To what extent can the answers to these questions shed light on turnout decline and why young people have become increasingly unlikely to vote? The following analysis attempts to provide some answer to these questions.

To examine how these factors are related to a person's likelihood of voting, we analyse the combined British election surveys, from 1979 to 2015, which gives us data on individuals' social characteristics and strength of attachment to political parties, and data on elections to do with the expected closeness of the race (Heath 2018). We can specify the model algebraically as:

$$Logit = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6,$$

where our dependent variable is the log odds of someone voting, and:

X_1 refers to an individual's sex (1 = male; 0 = female);
X_2 refers to an individual's age (1 = 35 years old and under; 0 = over 35 years old);
X_3 refers to an individual's level of education (1 = university educated; 0 = school only);
X_4 refers to an individual's strength of party identification: from 0 (no attachment) to 3 (strong attachment);

X_5 refers to an individual's perception of the policy difference between the parties: from 0 (no difference) to 1 (some difference) to 2 (big difference);
X_6 refers to the average gap in the opinion polls between the two main parties' share of the vote (measured in percentage points).

This model can be interpreted in much the same way as the multiple linear regression model, as long as we remember that the dependent variable refers to the log odds of voting, not *Y* or probabilities. We are not trying to predict the value of the dependent variable, but the likelihood (in terms of log odds) of the dependent variable taking one value (voting) rather than another (not voting).

Having specified our model, we state the null hypothesis for each independent variable, make a decision on whether we reject (or fail to reject) the null hypothesis, and then, if we reject, discuss the pattern of the association between the independent and dependent variable and whether or not it conforms with our theoretical expectations. These steps are exactly the same as for OLS regression, though the way we carry them out is slightly different.

The null hypothesis is that, for example, controlling for the other variables in the model, there is no association between a person's age and whether or not they voted. If the null hypothesis is true, then the partial regression coefficient (the log odds) for the age term will be zero. The significance test for this hypothesis is much the same as the other significance tests we have described earlier. We look to see if the coefficient is significantly different from zero by comparing the magnitude of the coefficient with its standard error. However, the way we do this in logistic regression is slightly different from how we did it in linear regression. Rather than calculate the test statistic, we calculate something called the z statistic (see Box 17.2).

The square of the z statistic is often known as the Wald statistic (reported in most statistical software packages), which has a chi-square distribution. The associated p value therefore indicates the probability of obtaining a Wald statistic of the observed size or greater if the null hypothesis were true. If the p value is less than 0.05, we can reject the null hypothesis at the 5 per cent level and go on to describe the pattern of the association. Recall that in a logistic regression we are in effect investigating how the likelihood of voting varies according to the range of predictors included in the model.

The results from this model are presented in Table 17.6. For the time being we will just focus on the significance of the parameter estimates and whether the association is positive

BOX 17.2 How to Calculate the Z Statistic

The formula for the z statistic is:

$$z = \frac{\hat{\beta}}{s\hat{e}\left(\hat{\beta}\right)}$$

where:

β is the estimated coefficient; and $s\hat{e}\left(\hat{\beta}\right)$ is the estimated standard error of the coefficient.

Table 17.6 Predicting turnout, 1979–2015 (logistic regression, parameter estimates)

	Log odds ratios	Standard error	Wald	Degrees of freedom (DF)	*P* svalue	Odds ratios
Sex: Female (REF)						
Male	–0.13*	0.040	10.4	1	0.001	0.88
Age: Over 35 (REF)						
35 and under	–0.54*	0.042	167.4	1	0.000	0.58
Education: No degree (REF)						
Degree	0.61*	0.074	67.9	1	0.000	1.83
Strength of party ID	0.59*	0.023	650.1	1	0.000	1.80
Policy difference	0.31*	0.026	137.4	1	0.000	1.36
Poll gap between parties	–0.03*	0.003	70.6	1	0.000	0.98
Constant	0.93	0.058	247.6	1	0.000	2.53

Notes: N = 22,004; chi square: 1,312; degrees of freedom: 6; * denotes *p* value <0.05;–2LL: 16748; Nagelkerke R-Square: 0.10.
Source: British Election Study 1979–2015, cumulative data set (Fieldhouse et al. 2016).

or negative. Table 17.6 reports both the log odds ratios and the odds ratios. It is usual practice to report the findings in terms of the log odds ratios, which have the statistical advantage of being centred on zero, making the direction of the association clearly evident (positive or negative). We find support for our expectation that social characteristics, such as sex, age, and education, matter for whether an individual will vote or not, even when we control for all the other variables in the model. For each of the social characteristic variables, we can reject the null hypothesis of no association, since the associated p values are all below 0.05. We indicate which coefficients are significant by an asterisk. From the direction of the coefficients, we can see that men tend to be less likely than women to vote (b = –0.13); people aged 35 and under tend to be less likely than people over 35 to vote (b = –0.54); and people who have been educated to degree level or above tend to be more likely to vote than those who have not (b = 0.61).

There is also a significant and positive association between turnout and strength of party attachment. We have treated strength of party attachment as an interval-level variable, and so the coefficient tells us that for every unit increase in the strength of attachment (on a 0 to 3 scale), the log odds of someone voting increase (b = 0.59). People with a strong party attachment therefore tend to be more likely to vote than people with only a weak attachment or no attachment at all. There is also evidence to support our hypothesis that the political context matters. There is a significant and positive association between turnout and perceptions of policy difference between the main parties (b = 0.31). Since the policy difference variable is also measured on an interval scale, this indicates that for every unit increase in the perception of policy difference, the log odds of someone voting become higher (b = 0.31). People who think that there is a great deal of policy difference between the parties are therefore more likely to vote than people who only think there is a small difference or no difference at all. Finally, there is a significant and negative association between the gap

between the main two parties' share of the vote in the opinion polls and turnout ($b = -0.03$). For every one percentage point difference between the Conservative and Labour share of the vote in opinion polls, the log odds of someone voting decrease by 0.03. This indicates that people are less likely to vote when a landslide victory is predicted for one party than when the race is expected to be very close.

Measures of fit

We are often interested in trying to establish 'how well' we have managed to explain our phenomenon of interest. Does our model of turnout provide a comprehensive explanation for why people vote or not? To try and answer this question, we often refer to what are known as measures of fit. As with OLS regression, there are a number of measures that can be used to describe how well the logistic regression model fits the data—or explains the observed variation in the data. However, they lack the intuitive appeal of R-square, they are somewhat more cumbersome to interpret, and there are disagreements about which measure is best. The logistic regression equivalent of R-square is known as pseudo R-square. The fact that it is prefaced with the word 'pseudo' indicates that it is not really the same as normal R-square, as it does not represent 'explained variation'. Rather, it refers to the proportional improvement in the fit to the data. To calculate the pseudo R-square we compare the model's fit to the data with the null model's fit to the data (when we only include the constant and do not include any independent variables), using something called the log likelihood (LL), or –2 log likelihood (–2LL) as it is often reported. There are different ways of doing this, and different measures of pseudo R-square are based on slightly different calculations, so it is always important to be clear about which measure is being used. In this example we report the Nagelkerke R-square. This is 0.10, indicating that the model improves the fit to the data by 10 per cent.

Interpreting the magnitude of coefficients

As already mentioned, it is difficult to meaningfully interpret the magnitude of the coefficients, and to compare which factors matter more than others. Although there is no direct equivalent to the standardized beta coefficient that we discussed with reference to OLS, we can get an idea about the relative magnitude of the different parameter estimates by comparing the size of the log odds ratio to its standard error. But a more intuitive way to illustrate the relative importance of different variables is to convert the log odds back into predicted probabilities, which are much easier to interpret. Although some statistical software packages, such as Stata, will calculate the predicted probabilities for you, it is not difficult to do by yourself in packages such as Excel (see Box 17.3). The change in predicted probability of voting (or change in predicted percentage) can then be neatly depicted in a graph, as shown in Figure 17.10, which illustrates the change in predicted turnout for a one-standard-deviation change in the independent variables, holding all other variables at their mean.

From Figure 17.10 we can clearly see that party identity—the strength of someone's attachment to a political party—has the largest impact on the predicted probability of voting. Holding all other variables at their mean, the predicted level of turnout increases by 4.5

BOX 17.3 How to Calculate Predicted Probabilities

From the results of our analysis, the predicted log odds of someone voting can be expressed by plugging the values of the different variables into our equation for the model, as follows:

$$Logit = 0.88 - 0.13 * X_1 - 0.62 * X_2 + 0.58 * X_3 + 0.59 * X_4 + 0.30 * X_5 - 0.03 * X_6$$

We can then use this equation to predict the log odds of someone voting, based on different values for each of the independent variables. The most straightforward way of doing this is simply to hold each value of *X* at its sample mean. Then, one by one, we can calculate the predicted log odds of voting for a one-standard-deviation increase in the value of each independent variable (remembering to return its value back to the mean once the calculation has been performed). For each variable, then, we have calculated the predicted log odds when all variables are held at their mean, and the predicted log odds when the variable in question has increased by one standard deviation. We can transform the predicted log odds into probabilities using the following function:

$$\breve{P} = \frac{1}{1 + \exp\left(-(a + b_1 X_1 + K\, K\, b_6 X_6\right)},$$

where exp is the inverse log function (or exponentiated function as it is often called).

percentage points for a one-standard-deviation increase in a person's strength of party attachment. The next most important variables are age and perceptions of policy difference. By contrast, the impact of sex is relatively weak (even though it is significant). From this we can see that how attached people are to political parties, and how much difference they think there is between them in policy terms are both strongly related to how likely someone is to vote. By contrast, civic resources, measured in terms of whether someone has been to university or not, is more weakly associated with turnout. This suggests that factors to do with parties and policy (at least in terms of how they are perceived) may be rather more important than factors to do with voters (in terms of their civic resources). People interested

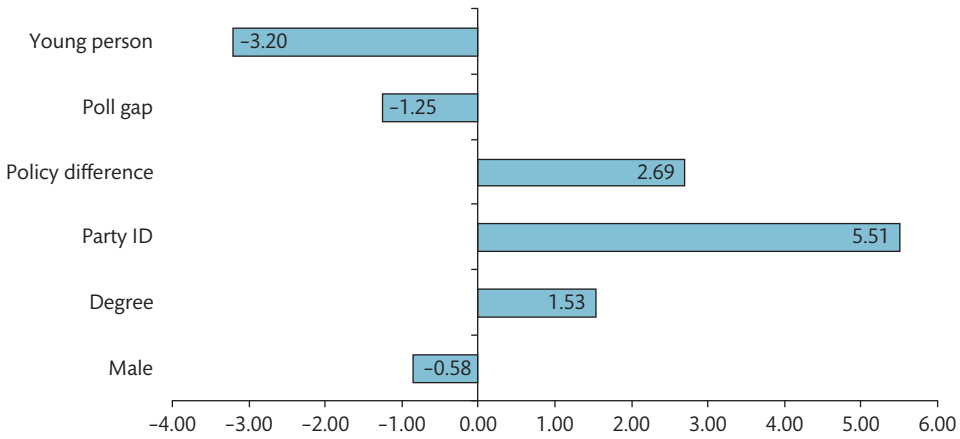


Figure 17.10 Change in predicted turnout for a one-standard-deviation change in the independent variables

Research in Action 17.3 Who Voted for Brexit?

The United Kingdom's vote to leave the European Union marked a watershed moment in the country's history. The result came as a shock to many, and in the immediate aftermath of the result political scientists sought to understand why people voted the way they did. One of the first such studies was carried out by Sara Hobolt (2016). Analysing survey data collected just before the referendum took place, Hobolt used logistic regression to examine which groups of voters were most likely to support Brexit. According to the long literature on Eurosceptism the so-called 'winners'of globalization—the young, well-educated professionals in urban centres—favour more open borders, immigration, and international cooperation, and so may be expected to support remaining in the EU, whereas the 'left behind'—the working class, less educated, and older people—oppose such openness, and so may be expected to support leaving the EU. Hobolt, and many subsequent studies, found that there were pronounced age and educational divides between people who supported 'Remain' and people who supported 'Leave'. In the multiple logistic regression models Hobolt shows that the impact of age and education on vote choices is highly significant, even controlling for lots of other demographic factors. In order to ease interpretation of the log odds, Hobolt shows the marginal effect of a one-standard-deviation change in each of the explanatory variables on the probability of voting Leave. This shows the strongest effects for education and age. Going from A-level education to an undergraduate degree reduces the probability of voting Leave by about ten percentage points, all other things being equal. Similarly, a 50-year-old is ten percentage points more likely to support Brexit compared to a 33-year-old voter. These are very substantial differences. However, as Hobolt notes, the overall model fit is modest (pseudo R-squared of 0.09) which indicates that demographic factors alone are not well-equipped to fully understand how people voted.

in increasing the level of turnout may therefore want to think about what parties stand for, and how this is communicated to the electorate. (See Research in Action Box 17.3.)

Interaction effects

Up to now we have assumed that each independent variable has the same impact on the dependent variable, regardless of the other factors, such as a person's age. So, for example, we have assumed that both old and young people respond to the policy difference between the parties in the same way—that is, that the impact of policy difference on turnout is the same for both groups. But we might think that the picture is more complicated than that. For example, we might think that young people are more likely to be influenced by the political context than older people, since they have not acquired the 'habit of voting'. This suggests that different groups of people respond to different factors in different ways. When we have strong theoretical reasons for suspecting this, we can fit an interaction term, which tests to see whether the impact of one variable varies significantly across levels of another variable.

Interaction terms can be difficult to interpret correctly, but they are frequently used in quantitative research and so are well worth understanding. When done appropriately, interaction terms can also be used in an incredibly revealing way to help shed light on complex empirical problems. They help us to examine heterogeneity, as they don't impose a one-size-fits-all explanation on political phenomena. Implicitly, we make arguments that involve interaction terms all the time. For example, Geddes (1999) states that no theory of democratic

transition explains all real-world variation. This is clearly true, but it does not mean that we have to throw up our hands and give up trying to explain transitions. Rather, we can try to develop explanations that take into account this diversity, so, for example, we can recognize that popular pressure from below does not influence transitions to democracy in all cases, but it tends to be of much stronger influence in single-party authoritarian regimes than in military regimes. There is thus an interaction between the impact of civil society on democratic transition and prior regime type.

We can illustrate the principles of how to specify and interpret interaction effects with the following example of electoral turnout. Various scholars have suggested that people's early political experiences have important long-term effects (e.g. Franklin 2004). Accordingly, it is often said that voting is a habit that is either acquired or not at a relatively young age. People's early experiences with elections shape their future behaviour. When there is plenty at stake, young people go out to vote and acquire the 'habit' of voting, which lasts into later life. However, when there is not much difference between the parties, and they are viewed as being all the same, young people do not go out to vote, and thus acquire the habit of not voting. The key point here, then, is that once the habit of voting (or not voting) has been acquired, older people should be relatively immune to the short-term vagaries of the political context. If they have acquired the habit of voting, they will turn out to vote more or less regardless of whether there are substantial differences between the parties or not. By contrast, younger people, who have not acquired the habit of voting, will be far more responsive to the perception of how different the political parties are. We can test this expectation by fitting an interaction term between age and policy difference. The interaction is illustrated in Figure 17.11, and shows that we estimate an additional parameter for the interaction between a person's age and their perception of policy difference. Here age is a moderator variable, and we examine whether the association between perception of policy difference and turnout varies according to whether someone is under 35 years old or not.

We can also specify the model algebraically, where the interaction term is defined as a multiplicative term between age and policy difference:

$$Logit = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_2X_5$$

where our dependent variable is the log odds of someone voting, and:

X_1 refers to an individual's sex (1 = male; 0 = female);
X_2 refers to an individual's age (1 = 35 years old and under; 0 = over 35 years old);
X_3 refers to an individual's level of education (1 = university educated; 0 = school only);
X_4 refers to an individual's strength of party identification: from 0 (no attachment) to 3 (strong attachment);

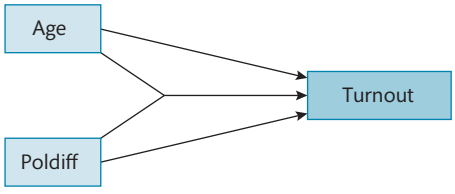


Figure 17.11 Multivariate model predicting turnout, with an interaction term

Table 17.7 Predicting turnout with an interaction term between age and policy difference (logistic regression, parameter estimates)

	Log odds	Standard error	Wald	Degrees of freedom (DF)	*P* value	Odds ratios
Male	–0.13*	0.04	10.2	1	0.001	0.88
Degree	0.60*	0.07	66.8	1	0.000	1.83
Strength of party identity	0.59*	0.02	652.4	1	0.000	1.80
Policy difference	0.26*	0.03	65.1	1	0.000	1.29
Gap in opinion polls	–0.03*	0.00	71.4	1	0.000	0.98
Young person	–0.73*	0.08	80.5	1	0.000	0.48
Young person by policy difference	0.15*	0.05	7.3	1	0.007	1.16
Constant	0.99	0.06	243.5	1	0.000	2.69

Notes: N = 22,004; chi square: 1,324; degrees of freedom: 7; *denotes p value <0.05;–2LL: 16737; Nagelkerke R-Square: 0.10.
Source: British Election Studies 1979–2015 (Fieldhouse et al. 2016).

X_5 refers to an individual's perception of the policy difference between the parties: from 0 (no difference) to 1 (some difference) to 2 (big difference);
X_6 refers to the average gap in the opinion polls between two main parties' share of the vote (measured in percentage points).

The results of this model are reported in Table 17.7. We can see that all the independent variables which are entered by themselves are significant and in the same direction as before. We will therefore limit our focus to the discussion of the interaction effect. Great care needs to be exercised when interpreting interaction effects, since they can often cause confusion. There are three parameters of interest. We have what are known as two parent terms—these are the variables that appear in the model on their own—and an interaction term, which is when these variables appear in combination with each other. We can see that each of the parent terms is significant and in the same direction as before (see Table 17.7). We can also see that the interaction term is significant ($p < 0.05$). This tells us that there is evidence that the association between policy difference and turnout is somewhat different for young people vs. older people.

Since older people are the reference category, the interaction terms denote how much bigger (or smaller) the magnitude of the policy difference coefficient is for young people. The interaction term between age and policy difference therefore tells us whether the association between policy difference and turnout is the same for young people (in which case the coefficient for the interaction term would be zero), stronger (in which case it would be positive), or weaker (in which case it would be negative). Since the interaction term is positive, we can conclude that the association between policy difference and turnout is stronger for young people than it is for older people. The estimated size of the policy difference coefficient for young people is therefore the size of the parent term coefficient ($b = 0.26$) plus the size of the interaction term coefficient ($b = 0.15$), which gives $b = 0.41$, whereas the

Table 17.8 Predicted probabilities

	Parties are all the same (0)	Big differences between the parties (2)
35 years old and under	0.73	0.86
Over 35 years old	0.85	0.90

Source: British Election Studies 1979–2015 (Fieldhouse et al. 2016).

estimated size of the policy difference coefficient for older people is just the size of the parent term coefficient ($b = 0.26$).

We can illustrate what this means more clearly by using this information to predict levels of turnout for young people and older people at different levels of policy difference. We do this in the same way as before, and hold all the other variables at their mean. We then calculate the predicted probability of voting for a young person when there is the perception of no difference between the parties (coded 0) and when there is the perception of big differences between the parties (coded 2). We do the same for older people. The results are presented in Table 17.8.

Table 17.8 shows that when there is a perception of big policy differences between the parties, people over 35 years old are just slightly more likely to vote (0.90) than people aged 35 or under (0.86). Or, put a different way, holding all other factors constant, the predicted difference in their turnout rates is just four percentage points. However, when there is a perception of no difference between the parties, the predicted probability of voting is much higher for older people (0.85) than it is for young people (0.73), even if both groups are less likely to vote than when there are big policy differences. In this instance, holding all other factors constant, the predicted difference in their turnout rates is now twelve percentage points.

We can also see that, for older people, the predicted level of turnout is five percentage points higher when they think there are big differences between the parties compared with when they think there aren't any differences at all. By contrast, for young people, the predicted level of turnout is thirteen percentage points higher. From this we can infer that whether the parties are all the same or not does not matter so much to older people—they are quite likely to vote in any event, perhaps because they have been socialized into voting from an earlier age when there was more at stake during elections. However, for young people, it makes more difference how different the parties are perceived to be—when they are very different, young people are almost as likely as older people to vote, but when they are much the same, then young people are far less likely to vote and are also much less likely to vote than older people.

The results from this simple model shed some light on recent discussions about turnout decline. First, we can see that important factors in deciding whether people vote or not are to do with their level of party identification, and the structure of the political context (see Heath 2007). Over the past few years party identification has declined and the ideological difference between the parties has narrowed. All other things being equal, this can lead to a substantial decline in levels of turnout. However, more interestingly perhaps, we can think that these changes have had a disproportionate effect on young people. Since young people are more influenced by the political context, the policy convergence between the two major parties had a greater impact on them than it had on older people, and so they

are correspondingly less likely to vote. Part of the reason, then, why turnout among young people has been so low in recent elections may be because the parties have become more similar to each other. It is not that young people have changed per se—they have always paid close attention to the contest—but because the parties have changed. And as they have become more similar, there is less incentive to vote.

Presenting and reporting data

Whatever the type of analysis that is being undertaken, it is important to present and report the information in a clear and accessible way. Tables must contain all the relevant information and be clear and easy to understand for the reader. In particular, it is important that they:

- have a clear title, with clear information about the dependent variable and the type of analysis used (logistic regression, OLS regression, etc.);
- indicate clearly the coefficient (unstandardized and/or standardized beta coefficients for OLS; log odds ratios and/or odds ratios for logistic regression);
- have clear labels for the independent variables (always give a self-explanatory label);
- indicate what the 'reference category' is for categorical variables;
- give information that enables the reader to judge the level of significance of the parameter estimates (asterisks and/or standard errors);
- report all parameter estimates (including non-significant ones);
- provide some indication of goodness of fit (R-square for OLS; pseudo R-square, chi square for logistic regression);
- report the source of the data;
- report the base (sample size) for the actual analysis.

When writing up the analysis, it is good practice to avoid causal language of 'effects', 'drivers', and so on and use language simply as if you were reporting patterns of association. Concentrate on whether the results confirm or refute your prior theoretical expectations, with reference to the direction of the parameter estimate, as well as on the level of significance. We have compared the sizes of the parameter estimates only in the vaguest and most general terms. If we want to provide a meaningful interpretation, we can calculate and present the predicted probabilities (for logistic regression) or the predicted values of y (for OLS regression).

Conclusions

Multivariate analysis is a powerful tool for exploring the relationship between different variables. It allows us to develop fuller accounts of why things happen, and to see whether these accounts are robust or not. Multivariate analysis also allows us to examine different types of relationship. We can use multivariate analysis to debunk spurious relationships, or to illustrate indirect causal mechanisms. We can also use it to specify interaction effects, and to examine

how the impact of one variable depends upon the level of another. These techniques are incredibly powerful, and can be used to good effect to explore a wide range of political phenomena. But, at the same time, the results from multivariate analysis should always be analysed with caution. Do the results really support the argument that the author is using them for? Is there a strong link between the theoretical expectations, and how this theory has been operationalized and tested? Have any potentially important variables been left out? How might the inclusion of these variables change the results? Have the variables been measured in a satisfactory way? These questions dominate all quantitative studies, and it is up to the researcher to demonstrate that they have taken steps to answer them as best they can.

Questions

1. What is the difference between a spurious relationship and an indirect relationship? Illustrate your answer with examples.
2. Why should we treat the results from bivariate analysis with caution? Why might the findings not stand up in a multivariate analysis?
3. What is the difference between causality and association?
4. How do we know if a relationship between two variables is causal or not?

Guide to Further Reading

Agresti, A. (2007), *An Introduction to Categorical Data Analysis* (New York: John Wiley & Sons).
Presents a non-technical introduction to topics on categorical data analysis, such as odds ratios and logistic regression.

Agresti, A. and B. Finlay (2009), *Statistical Methods for the Social Sciences*, 4th edition (Upper Saddle River, NJ: Pearson-Prentice Hall).
Covers the basics of statistical description and inference, as well as more advanced topics on regression methods, including multiple regression, ANOVA and repeated measures ANOVA, analysis of co-variance, logistic regression, and generalized linear models.

Carlson, J. and M. Hyde (2002), *Doing Empirical Political Research* (Boston, MA: Houghton Mifflin).
Provides an introduction to quantitative analysis and covers some more advanced topics, with lots of examples from political research.

de Vaus, D. (2002), *Surveys in Social Research* (London: Routledge).
A detailed and accessible introduction to quantitative methods, with easy-to-read chapters on how to carry out analysis and interpret the results.

Field, A. (2017), *Discovering Statistics Using SPSS*, 5th edition (London: Sage Publications).
Provides an easy-to-follow introduction to quantitative analysis and step-by-step instructions on how to use the SPSS statistical software package to do your own analysis.

Johnson, J. B., H. Reynolds, and J. Mycoff (2007), *Political Science Research Methods*, 6th edition (Washington, DC: CQ Press).
Provides an introduction to quantitative analysis and covers some more advanced topics, with lots of examples from political research.

Rowntree, D. (2000), *Statistics without Tears: An Introduction for Non-Mathematicians* (Harmondsworth: Penguin).
An easy-to-read introduction to statistical methods.

References

Abou-Chadi, T. and M. Wagner. (2019). 'The electoral appeal of party strategies in postindustrial societies: When can the mainstream left succeed?', *Journal of Politics* 81(4) (October): 1405–19.

Agresti, A. (2002), *Categorical Data Analysis* (New York: John Wiley & Sons).

Fieldhouse, E., J. Green, G. Evans, H. Schmitt, C. van der Eijk, J. Mellon, and C. Prosser (2016), 'British Election Study, 2015: Face-to-Face Post-Election Survey' [data collection]. UK Data Service SN: 7972. http://dx.doi.org/10.5255/UKDA-SN-7972-1

Fieldhouse, E., J. Green, G. Evans, H. Schmitt, C. van der Eijk, J. Mellon, and C. Prosser (2018). 'British Election Study, 2017: Face-to-Face Post-Election Survey' [data collection]. http://doi.org/10.5255/UKDA-SN-8418-1

Franklin, M. (2004), *Voter Turnout and the Dynamics of Electoral Competition in Established Democracies since 1945* (Cambridge: Cambridge University Press).

Geddes, B. (1999), 'What Do We Know about Democratization after Twenty Years?', *Annual Review of Political Science* 2: 115–44.

Heath, O. (2007), 'Explaining Turnout Decline in Britain, 1964–2005: Party Identification and the Political Context', *Political Behavior* 29(4): 493–516.

Heath, O. (2018), 'Policy Alienation, Social Alienation and Working Class Abstention in Britain, 1964–2010', *British Journal of Political Science* 48(4): 1053–73.

Hobolt, S. (2016), 'The Brexit Vote: A Divided Nation, a Divided Continent', *Journal of European Public Policy* 23(9): 1259–77

Letki, N. (2008), 'Does Diversity Erode Social Cohesion? Social Capital and Race in British Neighbourhoods', *Political Studies* 56(1): 99–126

Lipset, S. M. (1959), 'Some Social Requisites of Democracy: Economic Development and Political Legitimacy', *American Political Science Review* 53: 69–105.

Norris, P. (2015). 'Democracy Cross-National Data' Release 4.0, Fall. http://pippanorris.com/data

Putnam, R. D. (2007), '*E Pluribus Unum*: Diversity and Community in the Twenty-First Century: The 2006 Johan Skytte Prize Lecture', *Scandinavian Political Studies* 30(2): 137–74.

For additional material, including web links, case studies, examples of research and links to datasets, view our online resources:
www.oup.com/he/halperin3e

APPENDIX

Finding and Citing Sources for Political Research

I. Types of Sources

Print Materials

Many of these materials can now be accessed online. See Chapter 10.

Bibliographic and Reference Materials

Social Sciences Citation Index (***SSCI***) An index to works cited in current social and behavioural science literature. The SSCI provides access to current information and retrospective data from 1956 onwards. Fields covered include anthropology, history, library science, law, linguistics, philosophy, psychology, political science, public health, sociology, urban studies, and women's studies. The SSCI adds approximately 60,000 new cited references per week.

Arts & Humanities Citation Index (***A&HCI***) Provides access to current information and retrospective data from 1975 onwards. Fields covered include archaeology, architecture, art, Asian studies, folklore, history, language, linguistics, philosophy, radio/television/film, religion, and theatre. A&HCI adds approximately 15,250 new cited references per week. http://science.thomsonreuters.com/mjl/scope/scope_ssci/

WorldCat The world's most comprehensive bibliography. It covers virtually all the material catalogued by any of several thousand member libraries, including books, serials, archival materials, maps, visual materials, computer files, and internet resources. http://www.worldcat.org

Library of Congress Catalog (***LC***) The LC is arguably the world's most comprehensive library. Its web page includes links to the catalogues of several hundred other research libraries, primarily in North America, the UK, and Australia. http://catalog.loc.gov/

Library of Congress Classification Outline Useful for identifying the LC call numbers that you should use to browse to find a particular kind of material. http://lcweb.loc.gov/catdir/cpso/lcco/lcco.html

Guide to Historical Literature (***American Historical Association***) The basic principle of organization is geographic, with some reliance on topical and chronological subdivisions. The geographic sections include brief overviews of scholarly trends, and helpful annotations; some also include guides to reference works and bibliographies. http://www.worldcat.org/search?qt=worldcat_org_all&q=historical+literature

Book Review Digest Indexes reviews appearing in about a hundred periodicals; includes brief excerpts. http://www.ebscohost.com/public/book-review-digest/

Books in Print The online version includes full text reviews from *Choice*, *Library Journal*, *New York Times*, and other review sources. http://www.booksinprint2.com

International Bibliography of the Social Sciences Contains bibliographic information from over 2,600 journals published in a hundred countries and over 6,000 books per year, in the fields of anthropology, economics, political science, and sociology. Coverage is from 1951 to the present, with quarterly updates. http://www.ibss.ac.uk

Dissertation Abstracts Online Subject, title, and author guide to virtually every American dissertation accepted at an accredited institution since 1861; selected Master's theses have been included since 1962. In addition, since 1988, the database has included citations for dissertations from fifty British universities; and citations and abstracts from Section C, Worldwide Dissertations (formerly European Dissertations). http://library.dialog.com/bluesheets/html/bl0035.html

ProQuest Dissertations and Theses Database Citations to dissertation and theses from around the world from 1861 to the present day, together with 1.2 million full text dissertations that are available for download in PDF format. The database offers full text for most of the dissertations added since 1997 and strong retrospective full text coverage for older graduate works. http://www.proquest.com/

Handbooks, Yearbooks, and Encyclopaedias that provide information on various facets of the political world.

The Statesman's Yearbook Information on every country in the world, covering key historical events, politics, economics, trade, and infrastructure. http://www.statesmansyearbook.com/

Political Handbook of the World An overview of the political system and situation in countries around the world.

Europa World Yearbook Snapshots of the political, economic, and social conditions of countries worldwide.

World Encyclopaedia of Political Systems and Parties Explanation of executive, legislative, and judicial systems, political parties, and electoral systems around the world.

Public Opinion and Polling around the World: A Historical Encyclopedia World Encyclopaedia of Parliaments and Legislatures Country-by-country analysis of politics, parties, elections, and parliaments around the world.

The Encyclopaedia of Democracy Articles on the concepts of democracy, important figures, and country and regional assessments of the state of democracy in the world.

Encyclopaedia of World Political Systems A country-by-country survey of the political environment of a country and how that has affected governments.

Encyclopaedia of Nationalism The role of nationalist movements in politics, with entries for specific countries.

International Encyclopaedia of Elections Forms of elections and election processes.

Newspapers, News Services, News Digests, and News Magazines

Some of the sources, below, provide news summaries. Others are indexes to newspapers or to periodicals which report or analyse current events. Others tap various sources of current opinion and comment, or contain references to more general sources.

Newseum Allows you to find daily newspaper front pages from sixty-eight countries across the globe. http://www.newseum.org

World News Connection Provides English-language translations of material from thousands of media sources. Included are political speeches, television and radio broadcasts, and articles from newspapers, periodicals, and books. Online coverage: 1996 to the present. http://wnc.dialog.com/

For earlier coverage, see the *Foreign Broadcast Information Service. Daily Report*. Transcripts of foreign radio broadcasts, including news and comment.

Facts on File A weekly news digest (since 1941), indexed annually and with an attached atlas. http://www.fofweb.com

Keesing's Contemporary Archives A monthly record of world events (since 1931) with indexes. http://www.keesings.com

Annual Register of World Events http://annualregister.chadwyck.com

Newspaper Indexes A good index to a major newspaper provides an annual register of events. See, for instance, *The New York Times* Index or *The Times* (London) Index.

Reader's Guide to Periodical Literature Indexes general periodicals (magazines, including newsmagazines, but not scholarly journals). http://www.hwwilson.com/databases/Readersg.htm

Africa Digest Weekly record of African events (since 1961), with index.

Africa Research Bulletin (***Political and Cultural Series***) Indexed monthly summary of African political developments (since 1964).

Asian Recorder A weekly digest of Asian events (since 1955), with index.

The Middle East and North Africa. A survey and directory of the countries of the Middle East (since 1948) and North Africa (since 1964), with a listing of events.

BBC Worldwide Monitoring: International Reports The BBC's Monitoring Service systematically tracks broadcasts from 120 countries in fifty languages. The Summary of World Broadcasts (SWB) was a separate part of the service until the end of March 2001, when it was discontinued. After that date, the information previously carried in the SWB became part of the International Reports. Accessible through LexisNexis: http://www.lexisnexis.com

Academic and Trade Journals

JSTOR Reproduces the full image of over 117 scholarly journals, starting in each case with volume 1. The purpose is to provide an archival collection of core scholarly journals, not current access; most articles are at least two years old. Full text searching is available. Runs of additional titles are gradually being added to the database. http://www.jstor.org

Project MUSE Includes the full image of all articles in over a hundred scholarly journals, including all Johns Hopkins University Press journals. Full text searching, as well as subject searching, is available. http://muse.jhu.edu/

Web of Knowledge This source includes five databases covering virtually all fields of knowledge. Two of the most useful are the SSCI and the A&HCI. With these databases, you can discover networks of scholarly articles relevant to your particular research interests. Beginning with any article that you have found valuable, you can locate all subsequent journal articles that cite it. Registering for *Current Contents Connect*, which is part of the Web of Knowledge, allows you to set up a personal alerting profile for journals that you select. This profile automatically emails you the latest tables of contents for your chosen journals as they appear. Registering also enables you to set up citation alerts that notify you when future articles cite key articles you have found to be of special interest. *Current Contents Connect* includes a personal home page that displays all your selected titles and allows you to see the latest table of contents by clicking on a title. http://wokinfo.com/products_tools/products/

Social Sciences Full Text An index to English-language periodical articles. Subjects include anthropology, area studies, economics, international relations, law, political science, and sociology. Online coverage from February 1983, with abstracts from January 1994, and some full text articles. http://www.ebscohost.com/academic/social-sciences-full-test/

Ulrich's Periodicals Directory on the Web Provides detailed information on more than 300,000 periodicals of all types: academic and scholarly journals, e-journals, peer-reviewed titles, popular magazines, newspapers, newsletters, and more. http://ulrichsweb.serialssolutions.com/

Public Affairs Information Service Bulletin Indexes numerous periodicals in the field of public affairs and public policy: journals, pamphlets, government documents, some books. http://www.csa.com/factsheets/pais-set-c.php

Columbia International Affairs Online (CIAO) Allows search engine access to working papers from a wide range of think tanks and to tables of contents and article abstracts from the most recent issues of about thirty-five international affairs journals. Also contains information on forthcoming scholarly conferences in international affairs. http://www.ciaonet.org/

Fulltext Sources Online A systematic list of periodicals available online in full text, along with the names of the databases that contain them. Includes foreign- as well as English-language periodicals. http://www.fso-online.com

Historical Abstracts Selectively indexes over 2,000 journals in the field of history, with abstracts. Also indexes selected books. Online coverage from 1955. http://www.ebscohost.com/public/historical-abstracts

Humanities Full Text An index of English-language periodicals. Subjects include history, area studies, language and literature, philosophy, and religion. Online coverage: index from February 1984, abstracts from March 1994, and some full text from 1995.

International Political Science Abstracts Provides citations and abstracts for a worldwide selection of journal and yearbook articles in the fields of international relations, political science, and public administration. Online coverage: 1989 to present.

Social Sciences Index Indexes social science periodicals (scholarly journals).

Historical Abstracts Notes and abstracts of historical articles on the twentieth century. This journal began in 1955; in 1971, it spun off a second annual volume, *Twentieth-Century Abstracts*. Search it by country name.

International Political Science Abstracts Bimonthly abstracts of political science journals and yearbook articles. Find the name of the country you are researching in the index to be directed to abstracts of articles on that country.

Education Resources Information Center (ERIC) Provides access to more than 1.3 million bibliographic records of journal articles, books, research syntheses, conference papers, technical reports, policy papers, and other education-related materials, with hundreds of new records added multiple times per week. If available, links to full text are included. http://www.eric.ed.gov/

EServer Journals Collection Provides links to popular journals. Organized by subject. http://journals.eserver.org/

Government Reports and Legal Documents

PolicyFile: Public Policy Research and Analysis Indexes and abstracts of publications, covering a complete range of public policy research from such organizations as the American Enterprise Institute, Brookings Institution, Carnegie Endowment for International Peace, IMF, World Bank, CSIS, Rand Corporation, and many others. Where available, access to home pages and full text is provided. http://www.policyfile.com/

Public Affairs Information Service (PAIS) A selective listing, with abstracts, of journal articles, books, and documents. Subjects include economics, business, political science, sociology, demography, and international law and relations, with an emphasis on public-policy-orientated literature. Includes titles published in six languages. http://www.csa.com/factsheets/pais-set-c.php

LexisNexis Provides the full text of hundreds of publications, including law journals, wire services, country economic reports, government publications, magazines, newspapers, news digests, and industry newsletters and periodicals. http://www.lexisnexis.com

Internet-only Sources

Annotated List of Reference Websites (Library of Congress) Covers many fields of the social sciences and humanities. http://lcweb.loc.gov/rr/main/alcove9

Internet Scout Project Tracks useful websites and assesses their intellectual reliability. The Scout Report Archives is a searchable and browseable database of over seven years' worth of the Scout Report and subject-specific Scout Reports. It contains about 12,000 critical annotations of carefully selected internet sites and mailing lists. Each annotation analyses the site's general content, attribution (authors, etc.), currency, availability, accessibility, and presentation. http://scout.wisc.edu

Social Science Information Gateway A guide to a large number of online scholarly sites, screened by experts. Includes breakdowns by anthropology, economics, geography, politics, sociology, and other fields. Orientated primarily towards British scholarship. http://www.ariadne.ac.uk/issue2/sosig/

Weblogs/Blogs A type of interactive journal where writers post and readers respond. They vary widely in quality of information and validity of sources. Prestigious journalists and public figures may have blogs, which may be more credible than most.

Message boards, discussion lists, and chat rooms Some are useful and well-researched, others are not.

Multimedia The internet has a multitude of multimedia resources, including online broadcasts and news, radio, and television broadcasts, interactive talks, public meetings, images, audio files, and interactive websites.

Internet Data and Archive Resources

UK Data Archive provides access to the UK's largest collection of digital research data in the social sciences.

Data sets When you investigate a data set and its codebooks, consider the following questions: Who collected the data? What is the unit of analysis? How many cases are there in the data set? What was the sampling method? For how many variables was data collected in each case? What (dependent or independent) variables does it contain that are of particular interest to you? Can this data help to answer the question(s) you are -addressing?

Websites with links to lots of data sets

http://garnet.acns.fsu.edu/~phensel/data.html#index

http://www.psr.keele.ac.uk/data.htm

http://www.paulhensel.org/data.html

Political Resources on the Net Listings of political sites available on the internet, sorted by country, with links to parties, organizations, governments, media, and more from all around the world. http://www.politicalresources.net

Intute Social Sciences A selective catalogue of thousands of websites in the social sciences, hosted in the United Kingdom. Users can browse by topic and region or search by keyword. Each entry has been reviewed and annotated. The focus is on high-quality sites that provide information directly rather than just linking to other sites. A good resource for international social sciences data. http://www.intute.ac.uk/socialsciences

Internet Crossroads in Social Science Data Offers hundreds of annotated links to online data sources. Searchable by keyword or browseable by category, the site includes links to government and non-government sites concerned with domestic and international economics and labour, health, education, geography, history, politics, sociology, and demography. The site is maintained by the Data and Program Library Service at the University of Wisconsin, Madison. http://www.disc.wisc.edu/newcrossroads/index.asp

Inter-University Consortium for Political and Social Research (ICPSR, University of Michigan) A publicly available archive of datasets constructed by political scientists. http://www.icpsr.umich.edu/icpsrweb/ICPSR/

Data from official statistical agencies

These data are published on a yearly or quarterly basis by national and international statistical agencies.

National Statistical Offices and Data Archives such as the Council for European Social Science Data Archives (Cessda). http://www.cessda.org/

Official Statistics in Europe Web guide to socio-economic surveys. http://www.mzes.uni-mannheim.de/projekte/mikrodaten/drafts/index.html

Statistical yearbooks published by independent agencies and organizations, such as the statistical yearbooks from the *Encyclopaedia Britannica*, which provide data on governments, elections, economics, and demography; the yearbooks from SIPRI (Stockholm International Peace Research Institute) provide, among other things, data on military expenditures and warfare; the *World Handbook of Political and Social Indicators* (Yale University).

International statistical agencies such as the International Monetary Fund (IMF), the International Bank for Reconstruction and Development (IBRD), and the OECD (Organisation for Economic Co-operation and Development). The IMF provides international financial statistics, and trade statistics; the OECD provides historical statistics, employment statistics, and also economic surveys for individual countries; and the ILO (International Labour Organization) is a source of labour force statistics.

Websites of policy-making organizations There are many policy organizations, think tanks, private research institutes, and non-governmental organizations (NGOs) that produce relevant papers and data.

Bank for International Settlements http://www.bis.org

Bretton Woods http://www.brettonwoods.org

International Labour Organization (ILO) http://www.ilo.org/global/lang--en/index.htm

International Monetary Fund (IMF) *International Financial Statistics Online* (*IFS Online*) is the principal statistical publication of the IMF. Tables for each IMF member country include data on the country's exchange rates, IMF position, international liquidity, money and banking accounts, interest rates, prices, production, international transactions, government accounts, national accounts, and population. Selected series are published in area and world tables. Data may be downloaded as HTML, MS Excel, comma-delimited, or tab-delimited files. Some series begin with 1945. Updated monthly. http://www.imf.org

Organisation for Economic Co-operation and Development (OECD) includes many of the data sets available from the OECD. This is primarily useful for wealthy, developed countries, though comparison with other countries is sometimes included. http://www.oecd.org

Third World Network http://www.twnside.org.sg

United Nations (UN) http://www.un.org

United Nations Common Database (UNCDB): a broad-based statistical resource, it draws data from a wide variety of UN and UN-related (e.g. FAO, ILO, WHO) organizations.

UN Commission on International Trade Law. http://www.un.or.at/uncitral

UN Conference on Trade and Development (UNCTAD) http://www.unctad.org

World Bank http://www.worldbank.org

GDF Online (Global Development Finance Online): Global Development Finance (GDF) is the World Bank's annual report on external financing prospects for developing and transition countries. It tracks the yearly movement of international capital flows to developing countries, and analyses policy issues for developing countries. GDF provides statistical data for 137 countries. Data may be displayed as charts or maps and exported.

World Development Indicators includes nearly 800 statistical indicators related to social and economic development. It is organized in six sections: World View, People, Environment, Economy, States and Markets, and Global Links. The tables cover 152 economies and fourteen country groups—with basic indicators for a further fifty-five economies. Data can be downloaded for further manipulation. http://data.worldbank.org/data-catalog/world-development-indicators

World Economic Forum/Davos http://www.wef.org

World Trade Organization (WTO) http://www.wto.org

Information on National Institutions

Web Sites of National Parliaments (provided by the -Inter-Parliamentary Union), http://www.ipu.org/english/parlweb.htm

Websites for National Banks (provided by Bank for International Settlements). http://www.bis.org/cbanks.htm

Governments on the World Wide Web Comprehensive database of governmental institutions on the World Wide Web: parliaments, ministries, offices, law courts, embassies, city councils, public broadcasting corporations, central banks, multi-governmental institutions, etc. Includes also political parties. http://www.gksoft.com/govt/

The World Factbook (from the CIA) https://www.cia.gov/library/publications/the-world-factbook

The ESRC Data Archive The ESRC Data Archive is a resource centre whose holdings consist mainly of the data from past surveys. It holds almost 4,000 data sets, including those from many large and important government-produced surveys and censuses, as well as academic research and historical materials. http://www.esds.ac.uk/

II. Citing Sources

Cite While You Write (Cite While You Write™) is a feature of EndNote, a software package that facilitates proper citation.

The Internet Public Library's Style and Writing Guide Section Resources by subject, style, and writing guides. http://www.ipl.org/

Modern Language Association (MLA) has a brief overview of what MLA style is. The *MLA Style Crib Sheet* provides extensive descriptions of general MLA style notes, text, and block quotations, page formatting, MLA text citations, and the MLA works cited. http://www.mla.org/

The Online Writing Center (OWL) at Purdue An excellent overview, with information, example papers, and individual work; cited examples. http://owl.english.purdue.edu/owl/resource/747/08

College of DuPage Library, Citing Sources http://www.cod.edu/library/research/Citenet.htm

The Chicago Manual of Style Online: Chicago-Style Citation Quick Guide http://www.chicagomanualofstyle.org/tools_citationguide.html

Online! Citation Styles by A. Harnack and E. Kleppinger. http://www.bedfordstmartins.com/online/citex.html

Learning APA Style http://www.apastyle.org/learn

Glossary

Access: how the researcher enters the chosen **field** site and starts to do fieldwork.

Acquiescence bias: the tendency of some respondents to automatically agree with every question they are asked.

Aggregate data: describes data referring to aggregates or collectivities rather than individuals.

Association: two **variables** are associated when values of one variable vary systematically with values of the other variable.

Bandwagon effects: these are said to occur when people say they did something (for example, vote for a particular party), not because they necessarily did, but because they want to go along with the majority position.

Bar chart: a way of graphically displaying **nominal** or **ordinal** data.

Behaviouralism: the application of **positivism** and **empiricism** to the study and explanation of outcomes in Politics and International Relations; a school of thought concerned with the observable, tangible, objective facts of behavior, rather than with subjective phenomena such as thoughts, emotions, or impulses.

Bivariate analysis: analysis of the relationship between two **variables**.

Case: the **unit of analysis**.

Case selection: issues to do with how and why a particular **field** site or **case** is chosen for study.

Case study: the intensive study of one **case**.

Causality: a relationship between two **variables** in which changes in one variable *bring about* changes in another.

Causal mechanism: something that links a cause to its effect; that generates some type of 'necessary connection' between two events (the cause and the effect).

Central limit theorem: states that when **samples** are large (above about thirty), the **sampling distribution** will take the shape of a **normal distribution**.

Central tendency: a measure of central tendency is a way of summarizing the central value in a frequency distribution. See also **mean**, **median**, and **mode**.

Chi-square test: a statistical test used to test the **null hypothesis** that there is no relationship between two **variables**. Often used in conjunction with **cross-tabulations**.

Closed question: a question employed in a structured **interview** or **questionnaire** in which respondents are asked to select their answers from a limited range of options. Also called a 'fixed choice' question.

Closed setting: research settings with restricted **access** or which require permission to access.

Code: a short alphanumeric term that refers to the category of a **variable** and often the location of a text passage.

Coding: the process of identifying passages of text (or other meaningful phenomena, such as parts of images) and applying labels to them that indicate they are examples of some thematic idea. The coding process enables researchers quickly to retrieve and collect together all the text and other data that they have associated with some thematic idea so that they can be examined together, and different **cases** can be compared.

Comparative historical analysis: investigation that begins with a set of theoretical **propositions** and then uses multiple **cases** and time periods to illustrate, test, and refine them.

Comprehension problems: occur when, for a variety of reasons, the respondent does not fully understand the question that is being asked of them.

Concept-stretching: has to do with defining concepts in very broad terms in order to ensure that competing definitions or interpretations of the concept overlap with each other.

Confidence interval: range of values in which the true population value is thought to lie.

Confirmation bias: the tendency to pay attention only to those data that confirm the researcher's favoured explanation.

Confirmatory research: a type of research that involves testing hypotheses with evidence.

Constant: see **intercept**.

Construct validity: an approach used for assessing the **validity** of a measure which is based on the extent to which the measure is associated with other theoretically relevant factors.

Constructivism: an approach which maintains that reality does not exist as something independent of us, but is socially, and actively, *constructed*; it cannot be understood outside of the way humans interact.

Content analysis: analysis of the content of a text in order to uncover its meanings and intentions.

Content validity: an approach used for assessing the **validity** of a measure which is based on the extent to which the measure covers the full range of the concept, covering each of its different aspects.

Continuous variable: a continuous **variable** is measured on a continuous scale to varying degrees of precision, where within the range of the variable any value is possible—such as distance in metres or weight in kilos.

Contrast of contexts: A type of **comparative historical analysis** (comparable to the '**Most Different Systems Design**'; or MDSD) in which **cases** are selected to show how a similar feature—an institution or process general enough to be applicable across different contexts—produces different results in decidedly dissimilar contexts.

Control group: in experimental research, **subjects** in the control group do not receive the **intervention** and so are used as a point of comparison in order to gauge the effect of the intervention on the outcome of interest.

Correlation: a linear relationship between two **variables**.

Co-variance: a relationship between two **variables** such that one varies with the other.

Coverage error: occurs when the **sampling frame** from which the **sample** is drawn does not completely match the population of interest.

Covert research: when the researcher does not reveal their identity to the **informants**.

Covert participation: see **covert research**.

Crisp set: dichotomous **variables** that are used in Qualitative Comparative Analysis. In a crisp set analysis, a **case** is either in or out of a particular set.

Criterion validity: an approach used for assessing the **validity** of a measure which is based on how well the new measure of the concept relates to existing measures of the concept, or related concepts.

Critical discourse analysis: an approach that assumes that discursive power—control over a discourse—is a crucial constituent of social power, and a major means of reproducing dominance and hegemony; and that seeks to expose connections between language, power, and ideology.

Critical juncture: an event or founding moment which sets in motion a course of action, or a process or direction of development, that becomes difficult to reverse

Critical realism: A non-**empirical epistemology** that maintains the importance of identifying unobservable structures or elements in social life, but that also rejects the assumption that there are mind-independent facts.

Critical theory: a variety of theoretical approaches, including **constructivism**, feminist **theory**, **post-structuralism**, and post-colonialism that reject **positivism**, and the assumption that a distinction can be drawn between facts and values.

Cross-sectional comparative research: comparing two or more social settings or groups (usually countries) at one particular point in time.

Cross-sectional designs: involve making observations of a **sample**, or cross-section, of a population or phenomenon at a single point in time and analysing it carefully.

Cross-tabulation: a table that can be used to depict the relationship between two **variables**.

Crucial case: a type of **case study** which is based on the analysis of a **case** which is thought to be crucial for the confirmation or disconfirmation of a **theory**.

Data analysis: the processing, interpretation, and analysis of findings.

Data collection: A wide range of evidence gathering activities that enables researchers to subject rival causal hypotheses to **empirical** scrutiny including (but not limited to) observation, **participant observation**, experimentation, **interviews**, archival research, research gathered from secondary historical sources, ethnography, large-scale **surveys**, and the creation and modification of quantitative data sets.

Data reduction: the process of condensing (selecting, focusing, simplifying) the data that appear in written up **field** notes or transcriptions for the sake of manageability and so they can be made intelligible in terms of the issues being addressed.

Deduction: a process of reasoning that begins with a broad generalization or theoretical **proposition** and then moves to specific observations.

Deductive-nomological model: an explanatory model, also known as the 'covering law model', which holds that explanations of individual events or actions are derived from **laws** that express universal **empirical** associations.

Dependent variable: the outcome or phenomenon we are trying to explain.

Deviant case: a type of **case study** which is based on the analysis of a **case** which is known to deviate from established generalizations and does not fit the existing **theory**.

Discrete variable: a discrete **variable** can only take on certain fixed values within its range and cannot be subdivided.

Dispersion: the extent to which data values are clustered together or spread out. See also **standard deviation**.

Discourse analysis: a qualitative type of analysis that explores the ways in which discourses give legitimacy and meaning to social practices and institutions.

Double-barrelled question: a question that is really two questions in one.

Dummy variable: a dichotomous **variable** which can take one of two values (0 or 1).

Ecological validity: the extent to which the conditions simulated in the laboratory reflect real-life conditions.

Elite interviews: a method of obtaining, from political elites, information that might not otherwise be available to a researcher, or that can confirm the accuracy of information that has previously been collected from other sources.

Empirical: things that can be experienced through the five senses (of seeing, hearing, touching, smelling, tasting).

Empirical research: research based on first-hand gathering of data through, for instance, interviews or **surveys**.

Empirical theory: a **theory** that can be verified or disproved through observation or experiment.

Empiricism: a philosophical theory of knowledge which maintains that what we know of the world is limited to what can be observed. Knowledge is only that which originates in sensory experience: there is no a priori knowledge, no knowledge of reality that is acquired prior to sense experience.

Epistemology: a branch of philosophy which studies the nature, sources, and limits of knowledge.

Equivalence of meaning: has to do with whether the concept under investigation means the same thing to people in different contexts.

Event structure analysis (ESA): an analytic procedure that helps an analyst to 'unpack' an event—break it into its constituent parts—in order to develop a causal interpretation of what happened and why it happened as it did.

Experimental group: **subjects** in the experimental group which receive the **intervention**.

Expert interviews: a method of eliciting specific information from individuals with specialized knowledge or expertise on a particular issue.

Explanatory variable: see **independent variable**.

Exploratory research: research that operates as a guide to a process of discovery. It begins with a question or basic **proposition**, probes its plausibility against various types of data, and eventually generates a more concrete **hypothesis**, which can be more rigorously tested.

External criticism: involves an assessment of the authenticity of a source by evaluating its external characteristics, including when, where, why, by whom, and for whom the source was produced.

External validity: the extent to which findings can be generalized to other similar situations.

Falsifiability: a concept in the philosophy of science, popularized by Karl Popper, that for a **proposition** to be scientific, it must be logically possible to show, through observation or experiment, that it is false.

Face validity: an approach used for assessing the **validity** of a measure which is based on whether or not the measure intuitively seems like a good measure of the concept.

Faulty presupposition: see **presuming question**.

Field: the place where research is carried out.

Field experiments: experiments which take place in real-world settings.

Field notes: how the ethnographer records the data that they observe in the **field**.

Filter: a question used in **surveys** to filter out respondents for whom more detailed questions may not apply and who therefore should not be asked.

Focus group: a form of group interview organized to address a specific phenomenon of interest, a specific theme or topic, involving a group of people selected because they are related to, have experience of, or have beliefs about the topic or theme in question.

Foreshadowed problems: the questions or issues that the researcher intends to explore during their fieldwork.

Frequency distribution: displays the number of times each value of a **variable** occurs in a given **sample**.

Fuzzy set: multi-value **variables** that are used in Qualitative Comparative Analysis. In a fuzzy set analysis, a **case** can be either fully in or fully out of a particular set, or somewhere in between.

Gatekeeper: someone who controls **access** to **closed settings**.

Going native: when the researcher loses objective distance from the people that they are studying.

Grand theory: an attempt to construct a total theoretical system covering all aspects of social life, rather than an explanation of particular instances, societies, or phenomena.

Grounded theory: an inductive research strategy in which **theory** is produced through, and grounded in, data.

Histogram: a way of graphically displaying data for **interval** or continuous data.

Historical events research: research that examines particular events that either have an important impact on subsequent developments or provide an opportunity for testing the implications of a general **theory**.

Historical institutionalism: the study of how temporal processes and events influence the origin and transformation of institutions that govern political and economic relations.

Historical method: the techniques and guidelines historians use to find and evaluate evidence of the past and to construct accounts of the past based on this evidence.

Historical process research: research that focuses on how and why a series of events unfolded over some period of time.

Holism: the claim that the whole of something is distinct from and not directly explicable in terms of its parts. Social facts have to have social causes that are irreducible to facts about individuals.

Hypothesis: a tentative explanation—hunch, guess, assertion—that has not yet been tested against data collected in a concrete situation, but which it is possible to test by providing clear evidence for support or rejection.

Hypothetico-deductive model: an application of two operations: the formation of *hypotheses* and the *deduction* of consequences from them in order to arrive at beliefs which—although they are hypothetical—are well supported, through the way their deductive consequences fit in with our experiences and with our other well-supported beliefs (Føllesdal 1994: 234).

Independent variable: a factor we think may influence or cause changes in the **dependent variable**.

Indicator: **empirical** measure of an abstract concept.

Individualism: the claim that the basic units of society are individuals. Social phenomena are the combined results of individual actions.

Induction: reasoning from particular facts or observations to a general conclusion.

Inference: the process of drawing logical conclusions on the basis of premises assumed to be true.

Informants: the people who are being studied. Although this sounds a bit conspiratorial, the term implies something more than just 'respondents', since, whether unwittingly or not, informants provide information about the topic under study.

Informed Consent: the process in which researchers communicate information about a study to potential participants. Information delivered during this process includes, but is not limited to, the purpose of the study, the risks and benefits, and that participation is voluntary and can be discontinued at any time. Respondents should be made adequately and accurately aware of the type of information you want from them, why the information is being sought, what purpose it will be put to, how they are expected to participate in the study, and how it will directly or indirectly affect them.

Interaction effect: when the impact of one **independent variable** on the **dependent variable** varies according to the level of another independent **variable**.

Intercept: the point at which the **line of best fit** crosses the y-axis.

Intercoder reliability: the degree of **coding** consistency between two or more coders.

Internal criticism: the process of determining the **reliability** or accuracy of the information contained in the sources.

Internal validity: has to do with whether or not we are confident that a causal relationship really exists between two **variables**.

Interpretivism: the standpoint that recognizes the 'embedded' nature of the researcher, and the unique personal theoretical stances upon which each person bases their actions. It rejects the assertion that human behaviour can be codified in laws by identifying underlying regularities, and that society can be studied from a detached, objective, and impartial viewpoint by the researcher.

Interquartile range: the range of values of a **variable** within the middle 50 per cent of the distribution.

Interval: a **variable** where the values of the variable can be rank ordered, and the differences between two values are important and can be quantified and directly compared.

Intervening variable: a **variable** that helps to explain or account for the relationship between two other variables.

Intervention: in the data-gathering process, this occurs when the researcher manipulates or alters a **variable** of causal interest (such as what the subject is exposed to).

Interviewer effects: are said to happen in survey research when respondents answer identical questions in different ways depending upon the characteristics or identity of the person who asked them.

Interview guide: a set of questions that provide a list of talking points or areas to be covered by the interviewer in an interview.

Interview schedule: a list of questions to be asked of all interviewees.

Interviews: a method of **data collection** in which a researcher asks questions of participants.

Item non-response: occurs in survey research when the respondent does not provide a valid answer to a particular question.

Key informant: an informant that an ethnographer comes to rely upon as a valuable source of information and knowledge.

Laboratory experiments: experiments conducted in a facility that provides controlled conditions.

Large-N study: study of a large number of **cases** through statistical analysis.

Latent content: meanings that do not reside on the surface of communication and are therefore not easily observable.

Law: a statement of relationship (e.g. 'A leads to B') which is accepted as having been universally verified by observation.

Leading question: a question that encourages a specific response.

Level of measurement: how a **variable** is measured, and how the different values of the variable relate to each other. See also **nominal**, **ordinal**, **interval**.

Line graph: a way of graphically displaying data for repeated measures of the same **variable** over time.

Line of best fit: a straight line representing the best linear relationship between two **variables**. See also **regression line**.

listserv: a list management tool consisting of a set of email addresses for a group in which the sender can send one email and it will reach a variety of people.

Logical Positivism: an important extension of positivist thought that argues that logical reasoning and mathematics should also be treated as sources of knowledge, in addition to **empiricism**; that knowledge of the social world comes, not just through what can be observed (empiricism), but also through logical reasoning.

Log odds ratios: a statistical transformation of the **odds ratios**.

Longitudinal designs: explore changes or discern trends over time.

Macro-causal analysis: compares historical **cases** for the purpose of making causal **inferences** about macro-level structures and processes.

Manifest content: meanings that reside on the surface of communication and are therefore easily observable.

Mean: a measure of **central tendency** that is computed by adding the values for all the **cases** and dividing by the number of cases.

Measurement error: occurs when, for a variety of reasons, a respondent's answer to a question does not represent their true answer.

Median: a measure of **central tendency** that refers to the middle value in a rank-ordered set of **cases**.

Methodological holism: the claim that properties of a system as a whole cannot be deduced by the properties of its components alone. The system as a whole plays a role in how the parts behave.

Methodological individualism: the claim that explanations of social phenomena—classes, power, nations, or other social phenomena—must be reducible to the characteristics of individuals.

Methodology: the principles and procedures of inquiry used by researchers in a particular discipline.

Missing data: where a **variable** does not have a valid value the data is said to be missing.

Mixed-methods approach: a research approach that employs a variety of methods (see also **triangulation**).

Mode: a measure of **central tendency** that refers to the most common value of a **variable**.

Model specification: the way in which a **hypothesis** is tested and what **variables** are included in the analysis.

Most Different Systems Design (MDSD): method of comparison based on analysing **cases** which are very different from each other with respect to theoretically relevant **independent variables**.

Most Similar Systems Design (MSSD): method of comparison based on analysing **cases** which are very similar to each other with respect to theoretically relevant **independent variables**.

Multicolinearity: occurs in multiple **regression** when two **independent variables** are highly correlated with each other, making it difficult to disentangle the independent impact of either **variable**.

Multivariate analysis: analysis of the relationship between three or more **variables**.

N: denotes the number of **cases** in the analysis. Also referred to as **sample** size.

Narrative: a sequence of social actions and/or events that are linked step by step to preceding and subsequent ones and, generally, thematized, or emplotted.

Natural experiments: experiments in conditions which occur naturally, in so far as the researcher is not active in the data-gathering process.

Naturalism: the view that the social and behavioural sciences should have the same structure and logical characteristics as the natural sciences; that the primary goal of the social sciences—like that of the natural sciences—is to explain and predict social phenomena by means of **laws**; and that only that which can be observed, at least in principle, to be investigated scientifically (contrary to '**scientific realism**').

Nominal: a **variable** where the values or categories of the variable cannot be ranked or the differences between the categories quantified.

Non-probability sampling: sampling methods which do not use random selection.

Non-response: in survey research, refers to people who were eligible to take part in the survey, but for whatever reason did not. See **non-response bias**.

Non-response bias: is introduced into a **sample** when people who do not complete the survey are systematically different in some respect from those people who do answer.

Normal distribution: is a probability distribution. The graph of the distribution is symmetrical and bell-shaped. Sometimes known as the bell curve.

Normative research: research that addresses questions about what *should* or *ought to be*; that seeks, for instance, to propose or justify one value system over another.

Normative theory: **theory** concerned with questions about what is right and wrong, desirable or undesirable, just or unjust in society.

Null hypothesis: that in the population there is no association between two **variables**.

Objectivity: striving as far as possible to reduce or eliminate bias in the conduct of research.

Odds: the likelihood of one outcome occurring compared to the likelihood of an alternative outcome occurring.

Odds ratios: the **odds** of an event occurring in one group compared with another. An odds ratio of 1 indicates that the odds of a particular outcome are equal in both groups.

Omitted variable bias: when the causal significance of **variables** is exaggerated because important factors have been omitted from the analysis and therefore not taken into account.

Ontology: the study and the nature of what exists.

Open-ended question: a question where respondents are free to give their own answer and no response alternatives are provided.

Open setting: research settings with public **access** which do not require permission to enter.

Operationalization: process of turning abstract ideas into **empirical indicators**.

Ordinal variable: where the values or categories of the **variable** can be ranked, but the differences between the categories cannot be quantified or directly compared.

Outlier: a value that is markedly different to the other values in the distribution.

Overt research: when the researcher reveals their identity to the **informants**.

Paradigm: a conceptual scheme about which a community of scholars agree, and which defines the objects of investigation and the ways in which they are to be investigated.

Parallel demonstration of theory: a type of comparison that is concerned to show how a particular theoretical **proposition** plays out differently in different contexts.

Partial regression coefficients: a measure of the relationship between two **variables**, controlling for the effect of other variables.

Participant observation: type of fieldwork in which the researcher becomes a regular participant in the day-to-day activities of the people they are studying.

Path dependence: the notion that present institutions or cultural forms are the cumulative and selectively reproduced products of past social actions, which in turn provide the basis for future actions.

Population: the collection of units from which the **sample** is drawn and to which findings can be generalized.

Positive feedback: the notion that each step down a particular path reinforces the likelihood that future

changes will continue in the same direction because positive gains accrue to actors and make them reluctant to switch directions.

Positivism: the view that scientific knowledge of the social world is limited to what can be observed; and that we can explain and predict social phenomena by discovering empirical regularities, formulating law-like generalizations, and establishing causal relationships.

Post-structuralism: an approach concerned with how discourses are embedded in sets of power relations and supported by institutional and historical configurations; and, particularly, how the 'discursive practices' that lay claim to revealing knowledge are intertwined with the mechanisms of (political) power.

Presuming question: a survey question which assumes something that is not necessarily true.

Primary sources: materials that have been written or collected by actors who actually witnessed the events they describe. They provide direct or first-hand evidence about an event, object, or person, and show minimal or no mediation between the document/artefact and its creator.

Process tracing: a method that entails carefully tracing the sequence of events constituting a process (e.g. institutional persistence or change) in order to identify the underlying mechanism that connects hypothesized causes and outcomes.

Proposition: a statement which links two or more **variables**.

Purposive sampling: a **non-probability sampling** method which involves investigators using their own 'expert' judgement to select respondents whom they consider to be typical or representative of the population of interest.

Qualitative research: **empirical research** in which the researcher explores relationships using textual, rather than quantitative data. **Case study**, observation, and ethnography are considered forms of qualitative research. Results are not usually considered generalizable, but are often transferable.

Quantitative research: **empirical research** in which the researcher explores relationships using numeric data, Data that is numerical and can be 'counted' (e.g. responses to close-ended questions in a survey).

Questionnaire: pre-designed lists of **closed questions** designed to collect data from a large **sample** of respondents and to be completed by respondents themselves.

Questionnaire design: the way in which the survey questions are constructed and put together.

Question order effects: occur when the order in which questions are asked influences the answers that are given to them.

Question wording effects: occur in survey research when slight variations in the way in which questions are worded can have a significant impact on how people respond.

Quota sampling: is a **non-probability sampling** method which involves investigators selecting a certain number or quota of respondents to interview, according to the personal characteristics of the respondents, such as their age group, sex, and income group.

Random assignment: the experimental method for deciding which subject goes in which group.

Random digit dialling: is a sampling method for selecting people to take part in telephone **surveys** by generating telephone numbers at random.

Random sample: sampling method in which all units in the **sampling frame** have a known non-zero probability of being selected.

Range: the range of scores is the value of the smallest score subtracted from the largest score.

Rational choice theory: explains outcomes as the result of rational choices made by individuals within a given set of material and structural circumstances.

Reactivity: has to do with whether **subjects** are conscious of being studied and alter their behaviour accordingly.

Recall problems: error in survey research to do with memory failure, where respondents cannot remember whether or not they have done something or when they did it.

Regression: a statistical technique for describing the relationship between two or more **variables** where the **dependent variable** is **interval**-level.

Regression coefficient: see **slope coefficient**.

Regression line: see **line of best fit**.

Reliability: the extent to which a measurement produces the same results on repeated trials.

Replicability: the replicability of a study refers to whether or not it can be carried out in more or less the same way with a fresh group of **subjects** in another **sample** at a different point in time to check the **validity** and **reliability** of the research.

Replication: the practice of carrying out a similar data-collection exercise to see if the findings from the original study can be reproduced on an independent data set.

Research design: a logical structure of inquiry that specifies the sort of test or evidence that will convincingly confirm or disconfirm a **hypothesis**.

Response rate: the response rate of a survey refers to the percentage of people who completed the survey out of the total number of people selected for the survey.

Response variable: see **dependent variable**.

Retroduction: The interaction of **induction** and **deduction** in an evolving, dynamic process of discovery and **hypothesis** formation.

Sample: the **cases** which have been selected for analysis. A sample is a smaller subset of the population of interest. The way in which a sample is drawn affects the extent to which the sample can be used to make valid **inferences** about the population of interest.

Sampling distribution: the distribution of possible values for a statistic that we would expect to get if we drew repeated **samples** from a population.

Sampling error: the extent to which the **sample** estimate differs from the true population value.

Sampling frame: the list of units in the population from which the **sample** will be drawn.

Sampling variation: the extent to which the value of a **variable** varies in **samples** taken from the same population.

Scientific realism: the view that scientific research is based, not only on evidence that is visible in the world, but on unobservable facts, objects, or properties that exist independently of our direct knowledge of them, and that the goal of scientific inquiry is to describe and explain both observable and unobservable aspects of the world.

Secondary sources: materials that were produced sometime after an event happened, and so contain information that has been interpreted, commented on, analysed, or processed in some way.

Selection bias: the tendency to depend on accounts or select **cases** that fit with the researcher's arguments.

Semi-structured interviews: a form of interview in which the interviewer uses a combination of structured questions (to obtain factual information) and unstructured questions (to probe deeper into people's experiences).

Simple random sampling: is a probability sampling method where each member of the population has the same chance of being selected.

Single-N study: in-depth study of a single **case**.

Slope coefficient: derived from the **line of best fit**. Describes the amount the Y **variable** increases for a one-unit increase in the X variable. Also called the beta coefficient or the **regression coefficient**.

Small-N study: in-depth study of a small number of **cases**.

Snowball sampling: is a **non-probability sampling** method where respondents who have already taken part in the survey put forward their friends or acquaintances to also take part in the study.

Social desirability bias: occurs when a respondent provides an answer to a question which they think will present them in a good light, rather than one which reflects their true feelings.

Sparse categories: occurs when a table cell only contains a few **cases**. This can produce unstable results.

Speech act theory: an approach to the explanation of language premised on the idea that language is a medium orientated towards action and function.

Spurious association: occurs when an apparent association between two **variables** is caused by an unmeasured third variable. Spurious associations are not causal.

Standard deviation: a measure of **dispersion** of data points around the **mean**.

Standard error: the **standard deviation** of the **sampling distribution**.

Structured interviews: a form of interview that consists of a standardized set of closed and shorter or simpler questions that are asked in a standardized manner and sequence.

Subjects: the participants in the study.

Surveys: a method of **data collection** whose goal is to gain specific information about either a specific group or a representative **sample** of a particular group, and that includes at least one question which is either open-ended or closed-ended and employs an oral or written method for asking these questions.

Tails: values that lie at the extreme ends (i.e. tails) of a distribution.

Telescoping: occurs in survey research when respondents think that certain events from a long time ago happened more recently than they really did.

Temporality: in analysis, thinking explicitly about timing and sequence, and paying attention to the multiple forms of social time.

Theories of the middle-range: theories that, while they may involve abstractions, are constructed with reference to phenomena that are observable.

Theory: a **proposition** which has been elaborated and/or has withstood repeated testing.

Thick description: combines description and interpretation, and attempts to describe the meaning of an event and the motivations of the people involved.

Treatment: see **intervention**.

Triangulation: the use of a combination of three research methods in a study. An example of triangulation would be a study that incorporated **surveys**, **interviews**, and observations. (See also **mixed-methods approach**.)

Truth table: is used in Qualitative Comparative Analysis to display all the logical possible combinations of conditions and the outcome associated with each combination.

T test: a statistical test used to test the **null hypothesis** that there is no relationship between two **variables**.

Type I error: is said to occur when we reject a **null hypothesis** that is actually true.

Type II error: is said to occur when we fail to reject a **null hypothesis** that is actually false.

Unit of analysis: the entity that you are studying (e.g. the individuals or countries).

Univariate statistics: statistical measures used to summarize some characteristic of a single **variable**.

Unstructured interviews: a type of interview that uses open, and perhaps lengthier and more complex questions, which might vary in the way and order in which they are asked, and which allow for more in-depth probing and flexibility than **structured interviews**.

Validity: the extent to which our data measures the phenomenon we claim it measures. In the analysis of interview data, 'validity' refers to whether the conclusions being drawn from the data are credible, defensible, warranted, and whether they can withstand alternative explanations.

Variable: a property of an object that can take on different values.

Variance: the average amount of variability (spread) of a set of data.

Verifiability: the claim that the criterion for establishing truth claims, and thus the goal of social scientific inquiry, should be to verify statements or **propositions**.

Verification: the practice of making data available for secondary analysis so other scholars can re-analyse the data to see if they come to the same conclusions.

Volunteer sample: is a self-selecting **sample**.

Weighting the data: is a technique that is often used in survey research, which corrects for **non-response** or unequal selection probabilities to try to make the **sample** more representative of the population.

Reference

Føllesdal, D. (1994), 'Hermeneutics and the Hypothetical-Deductive Model', in M. Martin and L. C. McIntyre (eds), *Readings in the Philosophy of Social Science* (Cambridge, MA: MIT Press), 233–46.

Index

Tables, figures, and boxes are indicated by an italic t, f or b following the page number
The appendix (481–5) and glossary (487–96) will also provide information

A

acquiescence bias 295
African National Congress (ANC) 268–9
African-American industrial workers unionization 271–4
agendas, controlling 146
agent-structure problem 45, 46*(b)*, 56
  relationship between agents and structures 45
al Qaeda 371
Almond, on:, comparative research 232, 238
American Political Science Association, *Guide to Professional Ethics in Political Science (2012)* 181
analytical transparency 181
Angrist, on draft and the Vietnam War 213–14
annotation aids 382
Ansolabehere, on:
  'attack ads' 218
  negative campaigns 216–17
appendix, Finding and Citing Sources for Political Research 481–5
Arab-Israeli conflict 147
  US policy towards 269
archival sources
  in historical research 275
  primary and secondary sources 275–6
ArchiveGrid 276
ArchivesUSA 276
articulation 372
Arts & Humanities Citation Index (A&HCI) 100
association
  bivariate analysis and multivariate analysis 395–6
  and causality, bivariate analysis 425–7
  discourse and given context 368–9
  hypothetical example 442*(t)*
  patterns 422–3
  relationship between agents and structures 45
  sampling variation 432
  spurious 455–7, 462–3
  variables, correlation and co-variance 142

B

balanced questions 295
ballot design, and voting instructions 224
band-wagon effect 290
bar charts 400–1
barrios 347
behaviour
  'Heisenberg effect' 71
  'Rosenthal effect' 69–70
behavioural revolution 29
behaviouralism 29, 48, 77
beta coefficients 461*(t)*, 464, 468, 475
Bhavnani on electoral quotas for women 225
bias
  acquiescence bias 295
  confirmation bias 279
  interviews 311
  Nagel on 69
  publication bias 227
  'Rosenthal Effect' 69–70
  selection bias 194, 278
  social desirability bias 291
bibliographic/reference management, software packages 101
big data 201–4
bivariate analysis 395, 423–38
  categorical data, crosstabulation 438–48
  describing relationship 438–9
  ethnic diversity and civil war 446–7
  interval data 427–9
  and multivariate analysis 395–6
  ordinary least squares (OLS) regression 422
  Pearson's correlation 436, 437, 459
  process vs presentation types of answer 12
  significance testing with interval variables 431–2*(f)*
  statistical significance for cross-tabs 432
  strength of relationship 434–6
  test statistic 432
  using scatter plots 427, 429*(f)*
  using tables 440–1
Boolean logical comparison 247–9
Brexit 474*(b)*
British Election Study (2015) 197, 305, 440*(t)*, 440
British Election Study (2017) 397*(t)*, 400*(f)*, 441*(t)*, 441, 442*(b)*, 467*(t)*

C

calculations
  chi square 445
  interquartile ranges 406*(t)*, 407*(f)*
  line of best fit 429
  predicted probabilities 473*(b)*
  R-square 436*(b)*
  standard deviation 407*(b)*
  z statistic 470
canvassing and turnout 218*(t)*
capitalism
  and dependency theory 80–1
  and development theory 77
case selection, experimental research 214
case studies 17, 167, 234–7
  comparative research 167
  critical 235
  crucial cases 236
  democracy and development 427–8
  designs 167–5
  ethnography and 346
  generation of new theory 233
  outliers or deviant cases 401
  process tracing 237
  revelatory 235
  selection 167
  theory-confirming vs theory-infirming 233
  validity 192

categorical data, cross-tabulation 438–48
causal mechanisms, social world 41
causality, deterministic/probabalistic 247, 426
causation, causal mechanism
  defined 32, 41
  Hume on 32
  positivist concept 30–1
central limit theorem 410
central tendency 395, 402
checklist, Information Sheet and Consent Form 335*(b)*
chi square 441, 445
  degrees of freedom 471*(t)*
citation indexes 100
citizenship, Impact of Anti-Terrorism Policy on Citizenship 328*(b)*
Civil War, US South and 268–9
civil wars
  and ethnic diversity 446–7
  population pressures 215
classical positivism 30
clientelism, Benin, Wantchekon on 220
closed questions 294*(t)*
co-variance relationships 142
coder stability 385
coding 330, 333*(b)*
coefficient of reliability (CR) 385
coefficients
  interpreting magnitude 430, 434
  partial regression coefficients 461, 470
  regression coefficient 430, 470
  slope 430
cohort studies 167
Coleman, on:, religion and capitalism 44
Coleman's bathtub 44, 45*(f)*
colonialism 65
communist news items, wording effects 292*(t)*
comparative historical analysis (CHA) 263*(b)*–5, 274
  process tracing 269–70
  types 263–4
comparative and historical research 262*(b)*
Comparative Manifestos Project (CMP) 382
comparative research 167, 231–58
  Almond on 232, 238
  case selection 167
  case studies 167–5
  classic 238–9
  cross-sectional 262*(b)*, 274*(b)*
  data sources 171, 387
  descriptive function 237
  hypotheses 374–5
  large-N comparison 17, 251–5
  principles 232–4
  qualitative comparative analysis 247–51
  small-N issues 15, 217–18, 239
comprehension problems 303
Computer Assisted Qualitative Data Analysis Software (CAQDAS) 381
computer-assisted personal interviewing (CAPI) 303
Comté, on:
  classical positivism 30–1
  sociology 31
concepts 143–6
  anomaly concept 74, 83
  conceptualizing democracy 148
  formation 143–1
  power 146
  survey question validity 253
conceptual stretching 147
conceptual travelling 147
conceptualization
  from concepts to measurement 187*(f)*–90
  and operationalization 143
concordance 381
confidence interval 410, 411
  difference between two proportions 413–14
confirmation bias 279
confirmatory research 13*(b)*
  hypotheticodeductive model 12, 38, 50, 52
conflict
  ethnic heterogeneity and 446
  and 'groupness' 136, 141
  pressures 137*(f)*
confounding variables 456
  spurious relationships 454, 455–7, 462–3
consent 335*(b)*
constitutional durability 264
construct validity 287
constructivism 39, 53, 55, 65, 371
content analysis 377–86, 386
  data collection 374, 384
  dictionary-based 381
  latent content 388
  qualitative and quantitative 381
  steps 1-4, 377–80
  validity and reliability 373
  vs discourse analysis 387
contrast of contexts
  in CHA 263–5, 275
  logic 264
control groups 16
controlled interventions 219
correlation, and co-variance 142
corruption
  defined 190
  measuring 190–3
coverage error 296
covert participation 177
crisp set QCA 248
criterion validity 287
critical discourse analysis 367–71, 386
critical juncture 267
critical realism 47*(b)*
critical theory 65
  see also constructivism
Croatia
  independence declaration 57
  US, social interactions 58
cross-sectional comparative research 262*(b)*, 274*(b)*
cross-tabulation 438
  statistical significance 441–2
crucial cases 236
cultural theories, vs institutional theories 166

D

Dahl, *Polyarchy* 148
Dahl, on:
  democracy 189
  power 146
Dannreuther, on US policy towards the peace process 268–9
data
  big data 201–4
  Boolean analysis of data 247–9
  in historical research 275
  historical writing 275
  interval 427
  longitudinal data 167
  missing data 395, 396, 397
  observational 199–200
  primary/secondary data 195
  replication 197, 204
  sources
    archival, historical writing, and documents 275*(b)*
    'brute data' 48
  Twitter 202
  types 194–7
  verification 204, 205
  weighting 299*(b)*
data analysis 16
  Boolean 247–9
  categorical data, cross-tabulation 438–48
  coding 330

different processes of Data Analysis 329*(b)*
interviews 329
key terms 333*(b)*
presenting and reporting data 478
Data Analysis Software (CAQDAS) 381
data collection 14, 171–5
and data analysis 21*(t)*
discourse/content analysis 387
ethnographic research 171, 181
focus groups 130–3
from concepts to measurement 187*(f)*–90
from existing archives 275
inference 394
interview data 199–200
large-N comparison 251–5
multiple sources (triangulation) 19, 175
qualitative and quantitative data 12–16
questionnaires and surveys 171–9
reliability 192
replication 197, 204
research design 20–1*(t)*, 160
selection issues 193–4
textual analysis 359
transparency 204
validity 189–3
validity and reliability 373, 384
data gathering 171–9
data matrix, economic development and democracy 427–8*(t)*
data quality checklist 204–5
data reduction
interviews 329
process 332*(f)*, 333*(b)*
databases 100–2
JSTOR database 100
deduction 33, 35*(f)*, 35*(b)*, 38, 131–2
testing hypotheses 138, 139*(f)*
deductive reasoning 35*(b)*
deductive theory 131–2
deductive–nomological model, explanation 38
democracy
Dahl on 148
defining 148
development, OLS regression 437–8*(t)*
development theory 77
and ethnic diversity 446–7
and GDP 427–8, 430–1
questions on 100
democratic stability, Pakistan 241
dependency theory 78–81, 85
Geddes, on: 85
dependent variable 140–1*(b)*, 423
deprivation as a confounding variable 137*(f)*
deprivation hypothesis 134, 137*(f)*
descriptive hypothesis 128
descriptive inference 410–11
descriptive and inferential statistics 394–422
descriptive questions 170
research design 11, 108–9, 127, 128
deterministic causality 142, 143
deterrence theory 71
development, and democracy 427–8
development theory 77–8
democracy 427–8
Escobar on 367
Geddes on 85
Kuhn's model 77
modernization 85
QCA 248
underdevelopment 85
deviant cases 401
dictionary-based content analysis 381
Dion, on necessary and sufficient cause 243, 245
discourse analysis 386
articulation 372
critical (CDA) 367–71
'development discourse' 367
goal 368
hegemonic discourse 372–3
interpretive 278
text and context 369–70
validity and reliability 373
vs content analysis 387
discourse and social change, Fairclough on 372–3
discursive institutionalism 267
discursive power 367
'discursive practices', Renner on 371
dispersion, measures 405–6
distribution, normal 401*(f)*, 411*(f)*
distributive justice 67, 154
documents, computer-assisted qualitative data analysis software 381
Domestic Structures and European Integration (DOSEI) project 324
double-barrelled questions 290
duck vs rabbit 72*(f)*
Duverger's law 38

E

ecological validity 218
economic development and democracy, data matrix 427–8*(t)*
electoral behaviour *see* British Election Study
electoral fraud, in Armenia 223
electoral turnout *see* turnout
elite interviews 20, 322–4
empirical and normative questions, answers, methods 154
empirical research 13*(b)*
empirical theory, defined 132
empirical vs normative research 4–5
empirical vs normative theory 132
empiricism
classical tenet of positivism 30
and logic, basis of truth claims 33–40
environmental treaties, ratification 165
epistemology 7
defined 8, 28*(b)*
discourse analysis vs content analysis 387
holist/individualist 28, 40, 41
positivism, scientific realism and interpretivism 60*(b)*
error sum of squares (SSE) 436
errors
fallacies 112
fictional questions 113
measurement error 18, 288–9
proportional reduction in error (PRE) 436
proportions 412*(b)*
questionnaires and surveys 286*(f)*
of reasoning 122
sampling error 18, 301–2, 413
sources, in surveys 286*(f)*
tautology 113
Type I and II 432–3
Escobar, on 'development discourse' 367
ethical guidelines
experimental research 177–81
participants and 177, 226
ethical research 175–81
defined 175
Guide to Professional Ethics in Political Science (2012) 181
obligations of researchers 181*(b)*
participant observation 351*(b)*
researcher 180
sponsoring organization 180–1
surveys 302*(b)*
ethnic cleavage
Posner on 239
Zambia 239–40
ethnic conflict, Yugoslavia 51, 59*(b)*
ethnic conflict analysis 51–60
commitment problem 58
polarization 52–5, 59*(b)*

ETHNO diagram of causal relations 271–3*(f)*
ethnographer, role 350
ethnography
  case study and 346
  data collection 171, 181
  field notes 357–9
  foreshadowed problems 343
  identity of the researcher 355
  method 352–5
  recording observations 357
  Shifting Securities project 358
  theory and hypothesis testing 343–6
  see also participant observation
Europe
  Far Right 250
  Green parties 250
European Union
  Brexit 474*(b)*
  Domestic Structures and European Integration (DOSEI) project 324
Euroscepticism 332
event structure analysis 17, 265–6, 270–5
experimental groups 16
experimental research 150, 213–30
  case selection 214
  confidentiality 179
  design 213*(t)*
  design process 16, 164–5
  ethical guidelines 177–81
  field experiments 17, 218–20
  groups 16
  laboratory experiments 17, 214–18
  natural experiments 17, 164, 222–4
  obstacles to 225
  random assignment of participants 212
  reactivity 218
  Sixty-One-Million-Person Experiment on Social Media and Political Mobilization 218*(b)*
  strengths 212
  turnout 214, 219*(t)*
expert interviews 324, 333*(b)*
expert opinion 324
expert surveys, Freedom House and Polity IV 325
explanation, deductive-nomological model 38
explanatory hypothesis 129*(b)*
explanatory questions 11, 107–8*(b)*, 109, 112, 127, 170
explanatory variable 423
exploratory research 12, 13, 138
  selecting on dependent variable 423
'external criticism' 277

F

face validity 287
face-to-face interviews 303, 310, 318–19
facts vs values 32, 64
  normative and empirical theory in political research 132
  Weber on 69
Fairclough, on: discourse and social change 370*(f)*, 372–3
fallacies 112
false universalism 232, 233
falsifiability 39*(b)*
  Popper on 37
Far Right, Europe 250
faulty presuppositions 290
feminist theory 66
fictional questions 113
field experiments 17, 218–20
field notes 333*(b)*, 357–8
fieldwork, participant observation 339
filter questions 289
Finding and Citing Sources for Political Research 481–7
First World War 130–1
  Lenin on 150–2
focus groups
  ethical issues 334
  interviews 311–13, 325–30
  organizing 327–9
  participants 173, 311, 327
Foucault, on post-structuralist thinking 367
Frankel, on 'Heisenberg effect' 71
fraud, research fraud 205
Freedom House and Polity IV, expert surveys 325
Funtowicz and Ravetz on 'post-normal science' 81–2
fuzzy set QCA 248

G

game theory 20
gatekeepers 349
GDP, and democracy 427–8, 430–1
Geddes, on:
  case selection 17
  democratic transition 474–5
  dependency theory 85
general elections *see* British Election Study (2015); British Election Study (2017)
Geomans, War and Punishment (2000) 266*(b)*
Gershenkron's Economic Backwardness in Historical Perspective (1962) 267
globalization, dependency theory and 80–1
glossary 487–96
Google Flu trends 203
grand theory, vs theories of middle range 131
graphs and figures, quantitative analysis 400
Green and Gerber, on:
  controlled interventions 219
  obstacles to experimental research 225
  voting intention 218–19
Green parties, Western Europe 250
grounded theory 132
'groupness' 136, 141

H

hegemonic discourse 372–3
'Heisenberg effect' 71
Hempel, on:
  deductive-nomological model of explanation 38
  hypothetico-deductive model of confirmation 38
  laws 31
hermeneutics 48–9
Higgott, on Marxist theory 85
histograms 400, 401*(f)*
historical events research 262*(b)*, 274*(b)*
historical institutionalism 265, 266–7, 274*(b)*
historical process research 262*(b)*, 274*(b)*
historical research 260–83
  approaches 261–2
  convergence with social research 261
  temporality 261
historical writing as data 275
  primary and secondary sources 275–6
holism 39*(b)*
  and individualism 42*(b)*, 43–5
  and methodological 43
  methodological 43, 44*(f)*
holist ontology 43
Holsti, (1969), coefficient of reliability (CR) 385

honey-pot hypothesis 134, 137*(f)*
House of Commons *see* MPs
Hume, on:
causation 32
induction 32, 36
hypotheses 36–9, 142–3
association between discourse and given context 368–70
comparative research 374–5
components, outline 140
confirmatory 12, 38, 50, 52, 124
confirmed/falsified 20, 36–7
conflict, demographic and environmental pressures 137*(f)*
defined 131
deprivation 134, 137*(f)*
descriptive 128
explanatory 12, 127
falsifiable 13, 20, 36–7
formation 125
normative 128
null hypothesis 424, 432, 433–4
research questions and 129*(b)*, 138
resource curse hypothesis 137*(f)*
testing
induction 138, 139*(f)*
participant observation 342–3
variables 140*(b)*
vs propositions 131
hypothesis-generating research 138
hypothetical truth table, Ragin on 248–51, 249*(t)*
hypothetico-deductive model 39, 50, 52, 95, 125
of confirmation 38, 39
confirmatory research 12, 38, 50, 52, 124

I

ideologies 37
incommensurable knowledge 77
independent variable 140–1*(b)*, 423
multicolinearity 459, 460
indirect causality, intervening variables 456–7
individualism and holism 42*(b)*, 43, 43–5
induction 32, 131, 139*(f)*
methodological 43*(f)*
inducements, sampling 302
induction 35, 39*(b)*, 131–2
Hume on 32, 36
Popper on 35–7
rejecting, Popper on 36
testing hypotheses 138, 139*(f)*
inductive reasoning 34*(b)*
inductive theory 131–2
inference
descriptive inference 410–11
and scientific realism 46
Sinatra inference 237
informants, key 353
information sheet and consent form 335*(b)*
informed consent 335*(b)*
participants 177–8
institutional theories, vs cultural theories 166
institutionalism, historical 265, 266–7, 274*(b)*
intercept 430
intercoder reliability 192, 385
intermediate-N comparison
internal criticism 277
qualitative comparative analysis 247–51
'internal criticism' 277
internal validity 168
internal/external validity, case studies 196
international community, Blair on 373
international relations, vs politics 3–4
Internet
sources 119
surveys 303–4
interpellation, discourse analysis 372
interpretivism 5–7, 30, 47–52
defined 10
interquartile ranges 406*(t)*, 407*(f)*
positivism compared 60*(b)*
and scientific realism compared 60*(b)*
interquartile ranges 406*(t)*, 407*(f)*
interval data 427
interval variables 395, 399
intervening variables 140, 456–7
interviews
bias 311
computer-assisted personal 303
data analysis 329
data collection 199–200
data reduction 329
elite 20, 322–4
email 320
ending 319
expert 324, 333*(b)*
face-to-face 303, 310, 318–19
focus groups 311–13, 325–30
interview effect 314
interview guide 314–16*(b)*
interview mode 302
interviewer effect 303
online 303, 311, 312, 319–20
postal interviews 303
semi-structured 313, 314*(b)*, 333*(b)*
social movements 316*(b)*
structured 312, 314*(b)*, *333(b)*
telephone 306, 311, 312
unstructured 312, 313, 314*(b)*, 333*(b)*
video link and web 303, 321–2

J

JSTOR and Historical Abstracts 277

K

Kahl, on:
civil conflict 134–7*(f)*
'groupness' 136, 141
key words in context (KWIC), concordance 381
knowledge
forms of 8–9, 10, 28
incommensurable 77
Kuhn, on:
and Lakatos compared 86–8
nature of the social-scientific enterprise 82
paradigms 77
science as a social institution 86–8
scientific revolution 73–4, 76*(f)*

L

laboratory experiments 17, 215–18
advantages 215–16
Lakatos
impact on public attitudes 215
research programmes 85, 86–8
language, choice of, impact on public attitudes 215–18
large-N comparison
advantages 251
data sources 251–2
quantitative analysis 253
large-N studies 17, 251–5
Latin America
barrios 347
'New Left' 246
law/law-like generalization positivism 39
prediction of events 38
laws 32
rejecting 36
left-right self-placement
British adults 398–9*(t)*
Europe 404*(f)*, 405*(f)*, 409*(f)*, 418*(f)*
standard deviation 409*(f)*

Lenin, on causes of First World War 150–2
line of best fit 429, 434
linear relationship types 431*(f)*
Linguistic Inquiry and Word Count (LIWC) dictionary 381
listserv application 321–2
literature review 114–19
  vs survey 99
Little, on holist and individualist positions 43, 49
local government agencies 276
local public record offices 276
logical positivism 33, 35–6, 37
  Popper on 35–7
logistic regression 466–78
longitudinal data 167
longitudinal designs 262*(b)*
lottery winners, political attitudes 224

M

macro-causal analysis 263–5
macro-micro phenomena 43–5
macrosocial phenomena 44–6
Malawi, ethnic cleavage 239–40
manual and computer-assisted coding 381
Marxist theory, Higgott on 85
mean 403
  confidence intervals 417–18*(b)*
measurement error, indicators 288–9
measures of fit 472
media effects on public opinion 218
median 403
Members of Parliament 243–4, 353
  expenses scandal 414
  public satisfaction/dissatisfaction 415–16
metaphysical questions 113
methodological diversity 6, 8
methodological holism 43–4*(f)*, 43, 44*(f)*
methodological individualism 42*(b)*, 43, 44*(f)*
  and holism 43, 49
methodology 7–8, 28*(b)*, 43
micro level factors 44
Middle East peace process (MEPP) 268–70
migrants, refugees and 215–16
misreporting, deliberate 291
missing data 395, 396, 397
mixed methods approach 19
mode 403
model sum of squares (SSM) 435
moderator variables 457, 475
modernization theory 85
  Geddes on 80
Most Different Systems Design (MDSD) 241–2*(t)*
Most Similar Systems Design (MSSD) 239–44*(t)*
multicolinearity 459, 460
multinational corporations 80
multinational enterprises 262
multivariate analysis 395–6, 451–80
  advantages 451
  interaction effects 457
  missing data 396
  simple linear regression 427
  social trust 452–3
  statistical control 451–4
  using logistic regression 466–77
  using OLS regression 458–65
  variables 395

N

Nagel, on: bias 69
narratives
  competing interpretive narratives 278
  construction 270–1
national identity, shaping 371
National Inventory of Documentary Sources in the UK and Ireland (NIDS UK/Ireland) 276
national, state, or government archives 276
natural experiments 17, 164, 222–4
naturalism 31
  empiricism, laws (tenets of positivism) 31
nature of the social-scientific enterprise 82
necessary and sufficient cause, Dion on 243, 245
negative campaigning 216, 217*(t)*
  and likelihood of voting 217*(t)*
new industrializing countries (NICs) 252
'New Left', Latin America 246–7
newspapers 220
nomological model 38
non-governmental organizations (NGOs) 120*(b)*, 166, 347–9
non-probability sample design 298, 300
non-response 296
  bias 302
normal distribution 401*(f)*, 411*(f)*
'normal science' 76*(f)*, 81
normative arguments 153
normative hypothesis 128
normative institutionalism 267
normative political analysis 142, 152, 171
normative questions 11, 13*(b)*, 108, 111, 127, 170–1
normative questions, answers, methods 154
normative research design 152–3
normative theory 4, 30, 67, 132, 153
normative vs empirical research 4–5
null hypothesis 424, 432, 433–4
  degrees of freedom 471
  errors 432–3

O

objective criticism 355
objectivity and values 64–88
observation, separating facts and values 68–9
observational data 199–200
odds ratios, and log odds ratios 468, 472
OLS regression 430, 437–8*(t)*
  slope coefficient 430
  vs logistic regression 437–8*(t)*
online interviews 303, 311, 312, 319–20
ontology 5, 28*(b)*, 60
open-ended questions 294*(t)*
operationalization 149
  and conceptualization 143
ordinary least squares regression *see* OLS regression
outliers 401
  case studies 401

P

Pakistan, democratic stability 241
panel study 167
paradigms 74–8, 80
  anomalies 76, 80–4
  development theory 77
  in political research 77
parallel demonstration of theory 263–5, 274*(b)*
  contrast of contexts, and macro-causal analysis 263–4
partial regression coefficients 461, 470
participant observation 339–46
  choosing a research site 346–51
  ethical issues 351*(b)*
  methodology 344
  principles 339–46
  research site 349–51
  social movement groups (SMGs) 342

theory and hypothesis testing 342–3
thick description 343
participants
bias and 167
control and experimental groups 212
ethical issues 177, 226
ethnographic research 174
focus groups 173, 311, 327–8
informed consent 177–8
interviews 322, 326
reactivity 218, 350–1, 354
in research 177–9
harm to 179
participation
covert 177
vs turnout 49–50
party identification/attachment 235, 469
path dependence 267
patterns of association 424
line of best fit 425, 429
Peace Research Institute Oslo (PRIO) 194
Pearson's correlation 436
per cent agreement statistic 385
philosophy of social science 8–9
plurality system 39
policy analysis 111
political attitudes, lottery winners 224
political theory, positive 30
politics, vs international relations 3–4
Polity IV 325
polls, accuracy 306
Polyarchy (Dahl) 148
Popper, on:
empiricism and logic as basis of truth claims 35–7
logical positivism 33, 35–6
positive feedback 267
positive political theory 30
positivism
classical 30–3, 37
defined 9, 10, 13, 29–30
logical 33, 35–6
naturalism, empiricism, laws 31
pursuit of knowledge of the social world 32
research issues 5–7
three tenets 10
vs interpretivism 5–6, 10, 60*(b)*
vs scientific realism 40–2, 46, 60*(b)*
Posner, on ethnic cleavages 239
Post Office Address file 297, 304
post-colonialism 65–6
post-normal science 81
post-structuralism 367–8, 386
Foucault on 367
postal surveys 303
power concept 146–4
predictive questions 108, 110, 127, 170
prescriptive hypothesis 128
prescriptive questions 108, 110, 127, 170
presuming questions 289
primary and secondary sources, archives 275–7
probabalistic causality 142, 143
probabilities, predicted 473*(b)*
probability sample design 297, 298
process tracing 167, 237, 269–70, 274*(b)*
production transparency 181
programme eligibility randomization 223
proportional reduction in error (PRE) 436
proportions, confidence intervals 413*(f)*
proposition, defined 132
prospect theory 133, 264
Protestantism and capitalism 44
Weber on 44
pseudo R-square 472, 474
publication bias 227
purposive sampling 300

Q

qualitative comparative analysis (QCA) 243, 245, 247–9, 1018–22
qualitative content analysis 381, 384*(b)*
qualitative research, levels of measurement 197
quantitative analysis 396–400
descriptive and inferential statistics 394–422
large-N comparison 251–5
levels of measurement 197
tables 396–400
quantitative content analysis 381, 384*(b)*
programs 381
quantitative research, levels of measurement 197
quantitative vs qualitative research 6–8, 12–16
questionnaires
design 286–9
and surveys, data collection 171–9
questions
balanced 295
closed 294*(t)*
descriptive 11, 108–9, 127, 128, 170
double-barrelled 290
empirical and normative 154
explanatory 11, 107–8*(b)*, 109, 112, 127, 170
fictional 113
filter 289
hypotheses and 129*(b)*, 138
metaphysical 113
normative 11, 13*(b)*, 108, 111, 127, 154, 170–1
normative vs explanatory 12
open-ended 294*(t)*
order effects 291–4
predictive 108, 110, 127, 170
presuming 289
research design, formulating 8
survey question validity 253
validity 253
wording comprehension 288–9
wording effects 293–5
quota sampling 300

R

R-square
calculation 436*(b)*
explained/unexplained variation 435*(f)*–6
logistic regression equivalent (pseudo R-square) 472, 474
pseudo 472, 474
rabbit vs duck 72*(f)*
race and ethnicity 365
Ragin, on:
hypothetical truth table 249*(t)*
QCA 243, 245, 247–9
random allocation 223
random digit dialling 304
random probability sampling 297
randomization, eligibility 223
rational choice theory 30, 51, 55, 77, 88, 267
Rawls, on justice 153
reactivity
defined 218, 354
experimental research 218
participants 350–1, 354
recall problems (memory failure) 290
recontextualization 372, 373
reference/bibliographic management software packages 101
reflexivity 388
refugees, migrants and 215–16
regression, logistic regression 466–78
regression coefficient 430
partial 461, 470

regression line 430
sum of squares 430
regression line of best fit 434–5*(f)*
relationships, co-variance 142
reliability 192
coefficient of reliability (CR) 385
religions and idealogies 37
replicability 356
defined 356
replication, defined 197, 204
reproducibility 385
research, identity of the researcher 353
'research in action'
Do Voters Polarize When Radical Parties Enter Parliament? 409*(b)*
Does Diversity Erode Social Cohesion? 454*(b)*
Impact of Anti-Terrorism Policy on Citizenship 324*(b)*
National Narratives and Migration 375*(b)*
Sixty-One-Million-Person Experiment on Social Media and Political Mobilization 218*(b)*
The Decline of Social Democracy: When Can the Mainstream Left Succeed? 466*(b)*
'The Paradox of Political Persistence' 176*(b)*
Was There a Youthquake in the 2017 UK General Election? 442*(b)*
When and Why Do Wars End? 266*(b)*
Who Voted for Brexit? 474*(b)*
research components, outline 96*(b)*, 126*(b)*
Research Councils UK, Global Uncertainties programme 103*(b)*
research design 6, 13, 159–80
basic principles 161–3
comparative designs 167
components 60*(b)*
cross-sectional and longitudinal designs 165–6
cultural theories 167–8
data collection and data analysis 21*(t)*
drawing inferences 162
empirical vs normative 4–5
experimental process 16, 164–5
formulating questions 8
harm to participants 179
historical designs 168–9
key terms 13*(b)*, 163*(b)*
reliability 163
types 163–4, 163–5, 169–71
validity and 162
research fraud 205
research issues
politics vs international relations 3–4
positivism vs interpretivism 5–7
research positions/aims, pluralism 2
research presentation 125
research process
components 160*(b)*
Outline of Research Components 96*(b)*, 126*(b)*
overview 10–12
quantitative vs qualitative research 6–8
vs presentation 12
research questions 97–104
answering 11
formulating 94–6, 104–14
and hypotheses 129*(b)*
misconceptions 94, 94–5
Outline of Research Components 96*(b)*, 126*(b)*
research site
choice/access 319–22
gatekeepers 349
participant observation 349–51
types 320–1
research surveys, ethical issues 302*(b)*
research vase 104–5
resource curse hypothesis 134, 137*(f)*
resource scarcity/abundance 135
resource-related conflicts 135
response categories, continuous vs discrete 400
response format 296
response rates 305
response variable 423
retroduction 35*(b)*, 39*(b)*
revolutions, causes, MDSD and MSSD 241–2*(t)*
Robinson, on power 268
'Rosenthal effect' 69–70

S

sample design 296–8
sample mean 403
sampling distribution 410
sampling error 18, 296, 297*(b)*, 301–2
sampling frame 286*(f)*, 297
sampling, inducements 302
sampling variation 412
scatter plots 428
science
'as a social institution', Kuhn on 86–8
'normal science' 76*(f)*, 81
scientific realism
abduction and 46
inference and 46
interpretivism and positivism compared 60*(b)*
vs positivism 40–2, 46, 60*(b)*
scientific research programmes
key terms 83*(b)*
Lakatos on 86–8
scientific revolution, Kuhn on 73–4, 76*(f)*
Scottish elections, ballot design and voting instructions 224
secondary sources 275, 276–7
securitization issues 106
security, Shifting Securities ethnography project 358
selection bias 194, 278
data collection 194
significance *see* statistical significance
significance test (t-test) 432
with interval variables 431–2*(f)*
significance/insignificance, questions on research 99
silence, spiral of (theory) 344
simple linear regression 427
simple random sampling 298
Sinatra inference 237
Skocpol, States and Social Revolutions (1979) 265
Skocpol, on:
causal analysis 265
MDSD 242
social revolution 243–5, 263, 265, 267, 270
slope coefficient 430
small-N studies 17
snowball and volunteer sampling 300
social class 440–4
and UKIP 443
and voting, observed frequencies 441–4
social cohesion/trust 452–5
and civil war 446–7
and democracy 12 correlation coefficients 450
social construction 53
social desirability bias 291
social interactions, Croatia, US 58
Social Media and Political Mobilization 218*(b)*
social movement groups (SMGs) 342

social movements, interview guide 316*(b)*
social phenomena 27–8
Social Research Association, *Ethical Guidelines (2003)* 181
social revolution, Skocpol, on 243–5, 265, 267, 270
Social Sciences Citation Index (SSCI) 100
social trust 452–5
sociological (normative) institutionalism 267
sociology, invention 31
sources
  evaluation 277
  finding and citing 275–6
South Africa
  transition from white rule to democracy 268
  witchcraft 342
sparse categories 395
speech act theory 366, 386
spiral of silence theory 344
spurious association 455–7, 462–3
standard deviation 407*(b)*, 408*(t)*
standard error 411, 416, 417*(b)*
  difference between two proportions 417*(b)*
  regression coefficient 470
state, resource-related conflicts 135
state failure hypothesis 135, 137*(f)*
statistical analysis *see* bivariate analysis; multivariate analysis
statistical significance 416, 441
Structure of Scientific Revolution (Kuhn) 73–4, 76*(f)*
structure-agency debate 46*(b)*, 56
student performance, teachers' expectations 70, 221
sum of squares
  error (SSE) 436
  model (SSM) 436
  regression line 430
  total (TSS) 435
surveys 284–303
  concepts 253
  data collection 171–9
  ethical issues 302*(b)*
  expert surveys 325
  Internet 303–4
  platforms, PsyToolKit 215
  postal surveys 303
  question validity 253
  sources of errors 286*(f)*
  telephone surveys 304
  vs literature review 99
  wording of questions for 215–16*(t)*
Syria
  attitudes to people 216*(t)*
  war 215–16

T

t test 432
tables
  bivariate analysis 440–1
  cross-tabulation 438–45
  quantitative analysis 396–400
tails/outliers 401
Taliban, and al Qaeda 371
tautological errors 113
Taylor, on hermeneutics 48–9
teachers' expectations, and student performance 70, 221
telephone interviews 306, 311, 312
telephone surveys 304
telescoping 290
television news media, and crime 365
temporality 261, 274*(b)*
test statistic 433–4*(b)*
test-retest method 192
  reliability 192
textual analysis 364–90
  data collection 359
  discourse analysis 364–70
theories 123–56
  central limit theorem 410
  cultural theories 166
  deductive theory 131–2
  deliberative democracy 4
  dependency theory 78–81, 80–1, 85
    and third world 78–81, 85
  deterrence theory 71
  development theory 77, 85, 427–8
  empirical vs normative theory 132
  ethnography 343–6
  game theory 20
  grand theory 131
  institutional theories 166
  Marxist theory 85
  modernization theory 80, 85
  normative theory 4, 30, 67, 132, 153
Theory of Justice (Rawls) 153
theory-confirming vs theory-infirming case studies 233
third world, and dependency theory 78–81, 85
Tilly, on: causal mechanism 41
triangulation 19, 175
truth claims, empiricism and logic 35, 36, 39, 79
truth table 248–51
  hypothetical 249*(t)*
turnout
  2015 British General Election 440
  age and 232, 439, 440–1*(t)*, 442*(b)*
  attack ads 218
  canvassing and 218*(t)*
  decline in 217, 469
  early experiences with elections 475
  eligibility tests 188
  experimental design 214
  field experimentation 219*(t)*
  measurement 301
  non-response bias 302
  prediction 471*(t)*, 472, 473, 475*(t)*, 476*(f)*
  response bias 301
  vs participation 49–50

U

UKIP 443
US South, Civil War 268–9

V

validity 333*(b)*
  construct validity 287
  criterion validity 287
van Dijk, on:, discourse analysis 369–70
variables 140–1
  association 142
  confounding 454–7, 462–3
  dependent 423
  explanatory 423
  independent 140–1, 423
  interval 431–2*(f)*
  intervening 140, 456–7
  moderator 457, 475
  multicolinearity 459, 460
  relationships 142
  response 423
  and values 140*(b)*–1
verifiability, rejecting 36
verification of data 204
video link and web interviews 303, 321–2
Vietnam War, draft 213
volunteer sampling 298, 300
  2015 election 305
voting
  2015 British General Election 440*(t)*
  age and turnout 440
  negative campaigning 216, 217*(t)*
  see also turnout

W

Wald statistic 470
Wantchekon, on: clientelism in Benin 220
war
  and ethnic diversity 446–7
  First World War 130–1, 150–2
Washington, newspapers 220
Web *see* Internet
Weber, on:
  Coleman, religion, and capitalism 44
  distinction between facts and values 68–9
weighting the data 299*(b)*
women
  Bhavnani on electoral quotas for 225
  electoral quotas 225
  feminist theory 66
  in parliament 243–4
WordFish 383
wording of questions for surveys 215–16*(t)*
  migrants: refugees 215–16
WorldCat, primary source 276

Y

Yugoslavia, eruption of ethnic conflict 51, 59*(b)*